

# Optimal sampling design for global approximation of jump diffusion SDEs <sup>\*</sup>

Paweł Przybyłowicz

*AGH University of Science and Technology,  
Faculty of Applied Mathematics,  
Al. Mickiewicza 30, 30-059 Krakow, Poland,  
E-mail: pprzybyl@agh.edu.pl*

## Abstract

The paper deals with strong global approximation of SDEs driven by two independent processes: a nonhomogeneous Poisson process and a Wiener process. We assume that the jump and diffusion coefficients of the underlying SDE satisfy jump commutativity condition (see Chapter 6.3 in [21]). We establish the exact convergence rate of minimal errors that can be achieved by arbitrary algorithms based on a finite number of observations of the Poisson and Wiener processes. We consider classes of methods that use equidistant or nonequidistant sampling of the Poisson and Wiener processes. We provide a construction of optimal methods, based on the classical Milstein scheme, which asymptotically attain the established minimal errors. The analysis implies that methods based on nonequidistant mesh are more efficient than those based on the equidistant mesh.

**Key words:** nonhomogeneous Poisson process, Wiener process, jump commutativity condition, standard information, minimal strong error, asymptotically optimal algorithm

**Mathematics Subject Classification:** 68Q25, 65C30.

## 1 Introduction

We investigate the global approximation for the following jump diffusion stochastic differential equations (SDEs)

$$\begin{cases} dX(t) = a(t, X(t))dt + b(t, X(t))dW(t) + c(t, X(t-))dN(t), & t \in [0, T], \\ X(0) = x_0, \end{cases} \quad (1)$$

driven by two independent processes: a nonhomogeneous one-dimensional Poisson process  $N = \{N(t)\}_{t \in [0, T]}$  with intensity function  $\lambda = \lambda(t) > 0$  and one-dimensional Wiener process  $W$ . We assume, without the loss of generality, that  $x_0 \in \mathbb{R}$ . Jump diffusion SDEs (1) appear in various fields such as e.g. physics, biology, engineering and mathematical finance, see, for example, [1], [9], [29], [30] and pages 43-44 in [21]. We are interested in efficient algorithms that approximate whole trajectories of  $X$  and use only discrete values of the driving Poisson and Wiener processes.

---

<sup>\*</sup> This research was partly supported by the Polish NCN grant - decision No. DEC-2013/09/B/ST1/04275 and by AGH local grant.

Approximation of stochastic differential equations only driven by a Wiener process has been widely investigated in the literature. In that case, upper bounds on the error of defined methods were established, see, for example, [14]. Lower bounds were also investigated for the strong approximation in the Wiener (Gaussian) case, see, for example, [8], [12], [18]-[20] and [23]-[25].

In the jump diffusion case suitable approximation schemes were provided and upper bounds on their errors discussed, for example, in the monograph [21] and in the articles [3], [6], [9]-[11] and [16]. However, according to the author's best knowledge, till now there is only one paper that deals with asymptotic lower bounds and exact rate of convergence of the minimal errors for the global approximation of SDEs with jumps, see [26]. In that paper the author considered the pure jumps SDEs (1), i.e.,  $b \equiv 0$  and  $c = c(t)$ . We can also mention [4] where the authors investigated the optimal rate of convergence for the problem of approximating stochastic integrals of regular functions with respect to a homogeneous Poisson process. Here, we extend the approach used in [26] in order to cover more general SDEs of the form (1).

The purpose of this paper is to find lower bounds on the error and to define optimal methods solving (1). In the purely Gaussian case, similar question were considered, for example, in [12] and [18]. In order to study jump diffusion equations (1) driven by the Poisson and the Wiener processes a new technique is necessary. The main difference, comparing to the Gaussian case, is that we have to use some facts from the theory of stochastic integration with respect to càdlàg, square integrable martingales, see, for example, [15], [17], [22] and [29]. Moreover, we have to face the fact that when establishing the exact asymptotic constants the intensity of the process  $N$  depends on time. This problem does not appear in [26], where the intensity is constant. The another thing is we assume that the coefficients  $b$  and  $c$  satisfy the jump commutativity condition. This condition is widely described and discussed in, for example, Chapter 6.3 in [21]. Roughly speaking, it assures for the construction of the Itô-Taylor schemes that we do not need to know the exact location of the jump times of the Poisson process  $N$ . In this paper we widely use this condition when establishing asymptotic lower and upper bounds.

We consider three classes of approximation schemes denoted by  $\chi^{\text{eq}}$ ,  $\chi^{\text{noneq}*}$  and  $\chi^{\text{noneq}}$ , dependent on the sampling method for trajectories of the processes  $N$  and  $W$ . The class  $\chi^{\text{eq}}$  contains methods based on the equidistant discretization of  $[0, T]$ . Methods using the same (but not necessarily equidistant) evaluation points for  $N$  and  $W$  belong to a wider class  $\chi^{\text{noneq}*}$ . Methods that can use different, but also not necessarily equidistant, sampling point for the processes  $N$  and  $W$  belong to  $\chi^{\text{noneq}}$ . We have  $\chi^{\text{eq}} \subset \chi^{\text{noneq}*} \subset \chi^{\text{noneq}}$ .

The main result of the paper, Theorem 4.2, states that for fixed  $a, b, c, \lambda, x_0$  and in the case when the underlying SDE (1) is driven by two processes  $N$  and  $W$  (i.e.,  $b \not\equiv 0$  and  $c \not\equiv 0$ ) the following holds

$$\limsup_{n \rightarrow +\infty} n^{1/2} \cdot \inf_{\bar{X}_n \in \chi^{\text{noneq}}} \left( \mathbb{E} \|X - \bar{X}_n\|_{L^2([0, T])}^2 \right)^{1/2} \leq \frac{1}{\sqrt{6}} \int_0^T \left( \mathbb{E}(\mathcal{Y}(t)) \right)^{1/2} dt, \quad (2)$$

and

$$\liminf_{n \rightarrow +\infty} n^{1/2} \cdot \inf_{\bar{X}_n \in \chi^{\text{noneq}}} \left( \mathbb{E} \|X - \bar{X}_n\|_{L^2([0, T])}^2 \right)^{1/2} \geq \frac{1}{\sqrt{12}} \int_0^T \left( \mathbb{E}(\mathcal{Y}(t)) \right)^{1/2} dt, \quad (3)$$

where  $\mathcal{Y}(t) = |b(t, X(t))|^2 + \lambda(t) \cdot \mathbb{E}|c(t, X(t))|^2$ ,  $t \in [0, T]$ . In (2) and (3) the method  $\bar{X}_n$  uses at most  $n$  evaluations of  $N$  and  $W$ . By taking the infimum we mean that we choose mappings  $\{\bar{X}_n\}_{n \in \mathbb{N}}$  along with discretization points in the best possible way. For the subclass  $\chi^{\text{noneq}*}$  of  $\chi^{\text{noneq}}$  we have

$$\lim_{n \rightarrow +\infty} n^{1/2} \cdot \inf_{\bar{X}_n \in \chi^{\text{noneq}*}} \left( \mathbb{E} \|X - \bar{X}_n\|_{L^2([0, T])}^2 \right)^{1/2} = \frac{1}{\sqrt{6}} \int_0^T \left( \mathbb{E}(\mathcal{Y}(t)) \right)^{1/2} dt, \quad (4)$$

while in  $\chi^{\text{eq}}$  we have that

$$\lim_{n \rightarrow +\infty} n^{1/2} \cdot \inf_{\bar{X}_n \in \chi^{\text{eq}}} \left( \mathbb{E} \|X - \bar{X}_n\|_{L^2([0, T])}^2 \right)^{1/2} = \sqrt{\frac{T}{6}} \left( \int_0^T \mathbb{E}(\mathcal{Y}(t)) dt \right)^{1/2}. \quad (5)$$

In (5) the infimum means that we only choose mappings  $\{\bar{X}_n\}_{n \in \mathbb{N}}$  in the best possible way, while the discretization of  $[0, T]$  is fixed and uniform. As we can see, the order of convergence is  $n^{-1/2}$ , but the asymptotic constant in (5) may be considerably larger than that in (2), (3) and (4). In the class  $\chi^{\text{noneq}}$  we have a small gap between the upper and lower asymptotic constants. We conjecture that the exact rate of convergence of the minimal errors in  $\chi^{\text{noneq}}$  is the same as for  $\chi^{\text{noneq}*}$ . Note also that if  $b \equiv 0$  and  $c = c(t)$  then we arrive at results known from [26], while if  $c \equiv 0$  and  $b \not\equiv 0$  then, for the classes  $\chi^{\text{eq}}$  and  $\chi^{\text{noneq}*}$ , we restore the results known from [12], see Remark 4.2.

The asymptotically optimal scheme is defined by a piecewise linear interpolation of the classical Milstein steps, performed at suitably selected discretization points. The discretization points are chosen as quantiles of a distribution corresponding to a density  $\psi : [0, T] \rightarrow \mathbb{R}_+$ . It turns out that in the class  $\chi^{\text{noneq}*}$  the optimal density  $\psi_0$  is proportional to  $(\mathbb{E}(\mathcal{Y}(t)))^{1/2}$ . The main disadvantage of using such regular sampling is the need of using exact values of quantiles of  $(\mathbb{E}(\mathcal{Y}(t)))^{1/2}$  that might be hard to compute in general. In Section 4.1 we present the exact computation of sampling points in the linear case (Merton's model).

The paper is organized as follows. In Section 2 we give basic notions and definitions. Asymptotic lower bounds on the minimal errors are established in Section 3, while asymptotically optimal methods are defined in Section 4. We chose such order of presentation due to the fact that the technique used when proving the lower bounds in Section 3 suggests definitions of the optimal methods in Section 4. Finally, Appendix contains proofs of auxiliary results used in the paper.

## 2 Preliminaries

Let  $T > 0$  be a given real number. We denote  $\mathbb{N} = \{1, 2, \dots\}$  and  $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$ . Let  $(\Omega, \mathcal{F}, \mathbb{P})$  be a complete probability space. We consider on it two independent processes: a one-dimensional Wiener process  $W = \{W(t)\}_{t \in [0, T]}$  and a one-dimensional nonhomogeneous Poisson process  $N = \{N(t)\}_{t \in [0, T]}$  with continuous intensity function  $\lambda = \lambda(t) > 0$ . Let  $\{\mathcal{F}_t\}_{t \in [0, T]}$  denote the complete filtration generated by the driving processes  $N$  and  $W$ . We set  $m(t) = \int_0^t \lambda(s) ds$  and  $\Lambda(t, s) = m(t) - m(s)$  for  $t, s \in [0, T]$ . The process  $N$  has independent increments where the increment  $N(t) - N(s)$  has Poisson law with parameter  $\Lambda(t, s)$  and  $\mathbb{E}(N(t)) = m(t)$  for  $0 \leq s \leq t$ , see [7] or [21]. The compensated Poisson process  $\tilde{N} = \{\tilde{N}(t)\}_{t \in [0, T]}$  is defined as follows

$$\tilde{N}(t) = N(t) - m(t), \quad t \in [0, T], \quad (6)$$

which is a zero mean, square integrable  $\{\mathcal{F}_t\}_{t \in [0, T]}$ -martingale with càdlàg paths. For a random variable  $X : \Omega \rightarrow \mathbb{R}$  we write  $\|X\|_{L^q(\Omega)} = (\mathbb{E}|X|^q)^{1/q}$ ,  $q \in \{2, 4\}$ , and  $\|X\|_{L^2(\Omega)} = (\mathbb{E}(|X|^2 | \mathcal{G}))^{1/2}$ , where  $\mathcal{G}$  is a sub- $\sigma$ -field of  $\mathcal{F}$ . We say that a continuous function  $f : [0, T] \times \mathbb{R} \rightarrow \mathbb{R}$  belongs to  $C^{0,2}([0, T] \times \mathbb{R})$  if for  $j \in \{1, 2\}$  the partial derivatives  $\partial^j f / \partial y^j = \partial f^j(t, y) / \partial y^j$  exist and are continuous on  $(0, T) \times \mathbb{R}$ , and can be continuously extended to  $[0, T] \times \mathbb{R}$ . For a continuous function  $f : [0, T] \rightarrow \mathbb{R}$  its modulus of continuity is  $\bar{\omega}(f, \delta) = \sup_{t, s \in [0, T], |t-s| \leq \delta} |f(t) - f(s)|$ ,  $\delta \in [0, +\infty)$ . If  $Y = \{Y(t)\}_{t \in [0, T]}$  is a right-continuous process with left hand limits then we can define  $Y(t-) := \lim_{s \rightarrow t-} Y(s)$  for all  $t \in (0, T]$ . We have that  $Y(t-) = Y(t)$  if and only if  $Y$  is continuous at  $t$ . For the further properties of càdlàg mappings used in this paper see, for example, Chapter 2.9 in [1]. For  $f \in \{b, c\}$  we use the following notation

$$L_1 f(t, y) = b(t, y) \cdot \frac{\partial f}{\partial y}(t, y), \quad (7)$$

$$L_{-1} f(t, y) = f(t, y + c(t, y)) - f(t, y), \quad (t, y) \in [0, T] \times \mathbb{R}. \quad (8)$$

We impose the following assumption on the mappings  $a : [0, T] \times \mathbb{R} \rightarrow \mathbb{R}$ ,  $b : [0, T] \times \mathbb{R} \rightarrow \mathbb{R}$ ,  $c : [0, T] \times \mathbb{R} \rightarrow \mathbb{R}$  and on the intensity function  $\lambda : [0, T] \rightarrow \mathbb{R}_+$ :

(A)  $f \in C^{0,2}([0, T] \times \mathbb{R})$  for  $f \in \{a, b, c\}$ .

(B) There exists  $K > 0$  such that for  $f \in \{a, b, c\}$ , for all  $t, s \in [0, T]$  and all  $y, z \in \mathbb{R}$

$$(B1) \quad |f(t, y) - f(t, z)| \leq K|y - z|,$$

$$(B2) \quad |f(t, y) - f(s, y)| \leq K(1 + |y|)|t - s|,$$

$$(B3) \quad \left| \frac{\partial f}{\partial y}(t, y) - \frac{\partial f}{\partial y}(t, z) \right| \leq K|y - z|.$$

(C) There exists  $K > 0$  such that for  $f \in \{b, c\}$ , for all  $t \in [0, T]$  and all  $y, z \in \mathbb{R}$

$$|L_1 f(t, y) - L_1 f(t, z)| \leq K|y - z|.$$

(D) The diffusion and the jump coefficients satisfy the following *jump commutativity condition*

$$L_{-1}b(t, y) = L_1c(t, y), \quad (9)$$

for all  $(t, y) \in [0, T] \times \mathbb{R}$ . (We refer to Chapter 6.3 in [21] where the condition (9) is widely discussed.)

(E) The intensity  $\lambda : [0, T] \rightarrow \mathbb{R}_+$  is continuous in  $[0, T]$ .

The assumptions (B1) and (B2) imply for  $f \in \{a, b, c\}$  and all  $(t, y) \in [0, T] \times \mathbb{R}$  that

$$|f(t, y)| \leq K_1(1 + |y|), \quad (10)$$

where  $K_1 > 0$  depends only on  $f(0, 0)$ ,  $K$  and  $T$ . Moreover, by (B1) and (B3) we have for  $f \in \{a, b, c\}$  and all  $(t, y) \in [0, T] \times \mathbb{R}$  that

$$\left| \frac{\partial^j f}{\partial y^j}(t, y) \right| \leq K, \quad j = 1, 2. \quad (11)$$

From (B1), (10) and (11) we get for  $f \in \{b, c\}$  and all  $(t, y) \in [0, T] \times \mathbb{R}$  that

$$\max\{|L_{-1}f(t, y)|, |L_1f(t, y)|\} \leq K_2(1 + |y|), \quad (12)$$

where  $K_2 = KK_1$ .

Unless otherwise stated, all unspecified constants appearing in this paper may only depend on the constant  $K$  from the assumptions (B)-(C),  $x_0$ ,  $\|\lambda\|_\infty$ ,  $\|1/\lambda\|_\infty$ ,  $a(0, 0)$ ,  $b(0, 0)$ ,  $c(0, 0)$  and  $T$ . Moreover, the same symbol might be used to denote different constants.

The assumptions (A)-(E) are rather standard when comparing to those known from the literature concerning approximations of jump diffusion SDEs, see the comment before Theorem 6.1. Only in Section 4.1 we impose additional assumption on the coefficients which, in fact, turns out to be necessary in order to define an optimal sampling from a probabilistic density function.

For  $a$ ,  $b$ ,  $c$  and  $\lambda$  satisfying (B1), (B2) and (E) the equation (1) has a unique strong solution  $X = \{X(t)\}_{t \in [0, T]}$  that is adapted to  $\{\mathcal{F}_t\}_{t \in [0, T]}$  and has càdlàg paths, see [21], [22] or [30]. We have also the following moments estimates for the solution  $X$ , see, for example, [22] or [21].

**Lemma 2.1** *Let us assume that the mappings  $a$ ,  $b$ ,  $c$  and  $\lambda$  satisfy the assumptions (B1), (B2) and (E). Then there exist positive constants  $C_1$ ,  $C_2$  such that*

$$\left\| \sup_{t \in [0, T]} X(t) \right\|_{L^4(\Omega)} \leq C_1, \quad (13)$$

and for all  $t, s \in [0, T]$

$$\|X(t) - X(s)\|_{L^2(\Omega)} \leq C_2|t - s|^{1/2}. \quad (14)$$

The following result characterizes the local mean square smoothness of the solution  $X$  in the terms of the process  $\mathcal{Y} = \{\mathcal{Y}(t)\}_{t \in [0, T]}$  defined as follows

$$\mathcal{Y}(t) = |b(t, X(t))|^2 + \lambda(t) \cdot |c(t, X(t))|^2, \quad t \in [0, T]. \quad (15)$$

Of course  $\mathcal{Y}$  has càdlàg paths and it is adapted to  $\{\mathcal{F}_t\}_{t \in [0, T]}$ . (See Fact 6.1 in Appendix for the further properties of  $\mathcal{Y}$  used in the paper.)

**Proposition 2.1** *Let us assume that the mappings  $a$ ,  $b$ ,  $c$  and  $\lambda$  satisfy the assumptions (B1), (B2) and (E). Then for the solution  $X$  of (1) we have that for all  $t \in [0, T]$*

$$\lim_{h \rightarrow 0+} \frac{\|X(t+h) - X(t) \mid X(t)\|_{L^2(\Omega)}}{h^{1/2}} = (\mathcal{Y}(t))^{1/2}, \quad (16)$$

almost surely and, in particular,

$$\lim_{h \rightarrow 0+} \frac{\|X(t+h) - X(t)\|_{L^2(\Omega)}}{h^{1/2}} = \left(\mathbb{E}(\mathcal{Y}(t))\right)^{1/2}. \quad (17)$$

**Proof.** See the Appendix. ■

By Proposition 2.1 the square root of  $\mathcal{Y}$  can be interpreted as a conditional Hölder constant of  $X$ . This local smoothness will reflect in the exact rate of convergence of minimal errors established in Section 4. A result similar to Proposition 2.1 for SDEs driven by a multiplicative Wiener process has been obtained in [12], while for SDEs driven by an additive fractional Brownian motion with the Hurst parameter  $H \in (0, 1)$  has been shown in Proposition 1 in [19].

The problem considered in the paper is to find an *optimal strong global approximation* of the solution  $X = \{X(t)\}_{t \in [0, T]}$  of (1). For any fixed  $(a, b, c, \lambda, x_0)$  an approximation of  $X = X(a, b, c, \lambda, x_0)$  is given by a method  $\bar{X} = \bar{X}(a, b, c, \lambda, x_0)$ . The method computes the approximation by using some information about the functions  $a$ ,  $b$ ,  $c$  and  $\lambda$ , the Poisson process  $N$  and the Wiener process  $W$ . We consider methods that are based on a finite number of observations of trajectories of the driving processes  $N$  and  $W$  at suitably chosen points from the interval  $[0, T]$ . The cost of the method is measured by the total number of evaluations of the processes  $N$  and  $W$ .

We fix  $(a, b, c, \lambda, x_0)$  and we consider the corresponding equation (1). Any approximation method  $\bar{X} = \{\bar{X}_n\}_{n \in \mathbb{N}}$  is defined by three sequences  $\bar{\varphi} = \{\varphi_n\}_{n \in \mathbb{N}}$ ,  $\bar{\Delta}^Z = \{\Delta_n^Z\}_{n \in \mathbb{N}}$ ,  $Z \in \{N, W\}$ , where

$$\varphi_n : \mathbb{R}^{2n} \rightarrow L^2([0, T]), \quad (18)$$

is a measurable mapping and

$$\Delta_n^Z = \{t_{0,n}^Z, t_{1,n}^Z, \dots, t_{n,n}^Z\}, \quad (19)$$

is a partition of  $[0, T]$  with

$$0 = t_{0,n}^Z < t_{1,n}^Z < \dots < t_{n,n}^Z = T, \quad (20)$$

for  $Z \in \{N, W\}$ . We have that  $\{0, T\} \subset \Delta_n^N \cap \Delta_n^W$  for all  $n$  and, in particular, we might have  $\Delta_n^N \cap \Delta_n^W = \emptyset$  for some  $n$ . The sequences  $\bar{\Delta}^N$ ,  $\bar{\Delta}^W$  provide (not necessary equidistant)

discretizations of  $[0, T]$  used by  $N$  and  $W$ , respectively. Mostly, in the literature, we have that  $\bar{\Delta}^N = \bar{\Delta}^W$  see, for example, Chapter 6 in [21]. Here, mainly for the lower bound, we allow more general discretization. By

$$\bar{\mathcal{N}}(N, W) = \{\mathcal{N}_n(N, W)\}_{n \in \mathbb{N}}, \quad (21)$$

we denote a sequence of vectors  $\mathcal{N}_n$  of size  $2n$ , which provides *standard information* with  $n$  evaluations of the Poisson process and  $n$  evaluations of the Wiener process at the discrete points from  $\Delta_n^N \cup \Delta_n^W$ , i.e.,

$$\begin{aligned} \mathcal{N}_n(N, W) &= \mathcal{N}_n(N) \oplus \mathcal{N}_n(W) \\ &:= [N(t_{1,n}^N), N(t_{2,n}^N), \dots, N(t_{n,n}^N), W(t_{1,n}^W), W(t_{2,n}^W), \dots, W(t_{n,n}^W)], \end{aligned} \quad (22)$$

where  $\mathcal{N}_n(Z) = [Z(t_{1,n}^Z), Z(t_{2,n}^Z), \dots, Z(t_{n,n}^Z)]$  for  $Z \in \{N, W\}$ . Recall that  $N(0) = W(0) = 0$ . In particular, the sequences  $\bar{\varphi}$ ,  $\bar{\Delta}$  may depend on functions  $a$ ,  $b$ ,  $c$ ,  $\lambda$  and on  $x_0$  but not on trajectories of the processes  $N$  and  $W$ . (Information (22) uses the same evaluation points for all trajectories of the Poisson and Wiener processes.) Therefore, information (21) about the processes  $N$  and  $W$  is *nonadaptive*. Moreover, since  $\mathcal{N}_n(N, W)$  does not have to be contained in  $\mathcal{N}_{n+1}(N, W)$ , the information (22) is called *nonexpanding*, see [24]. We stress that our model of computation covers the regular strong Taylor approximations and it excludes the jump-adapted time discretizations, since we do not assume the knowledge of the jump times for  $N$  (see Chapters 6 and 8 in [21]). This restriction reflects our assumption that only nonadaptive standard information is available for the process  $N$ .

After computing the information  $\mathcal{N}_n(N, W)$ , we apply the mapping  $\varphi_n$  in order to obtain the  $n$ th approximation  $\bar{X}_n = \{\bar{X}_n(t)\}_{t \in [0, T]}$  in the following way

$$\bar{X}_n = \varphi_n(\mathcal{N}_n(N, W)). \quad (23)$$

The  $n$ th cost of the method  $\bar{X}$  is the total number of evaluations of  $N$  and  $W$  used by the  $n$ th approximation  $\bar{X}_n$ , defined as follows

$$\text{cost}_n(\bar{X}) = \begin{cases} 2n, & \text{if } b \not\equiv 0 \text{ and } c \not\equiv 0, \\ n, & \text{if } (b \not\equiv 0 \text{ and } c \equiv 0) \text{ or } (b \equiv 0 \text{ and } c \not\equiv 0), \\ 0, & \text{if } b \equiv 0 \text{ and } c \equiv 0. \end{cases} \quad (24)$$

(If  $b \equiv 0$  then we take formally  $\mathcal{N}_n(W)$  to be a zero vector and the sequence  $\bar{\Delta}^W$  can be arbitrary; we use analogous convention in the case when  $c \equiv 0$ .) The set of all methods  $\bar{X} = \{\bar{X}_n\}_{n \in \mathbb{N}}$ , defined as above, is denoted by  $\chi^{\text{noneq}}$ . Moreover, we consider the following subclasses of  $\chi^{\text{noneq}}$

$$\chi^{\text{noneq}*} = \{\bar{X} \in \chi^{\text{noneq}} \mid \exists_{n_0^* = n_0^*(\bar{X})} : \forall_{n \geq n_0^*} \Delta_n^N = \Delta_n^W\}, \quad (25)$$

and

$$\chi^{\text{eq}} = \{\bar{X} \in \chi^{\text{noneq}} \mid \exists_{n_0^* = n_0^*(\bar{X})} : \forall_{n \geq n_0^*} \Delta_n^N = \Delta_n^W = \{iT/n : i = 0, 1, \dots, n\}\}. \quad (26)$$

Methods based on the sequence of equidistant discretizations (19) belong to the class  $\chi^{\text{eq}}$  while to the class  $\chi^{\text{noneq}*}$  belong methods that evaluates  $N$  and  $W$  at the same, possibly nonuniform,

sampling points. We have that  $\chi^{\text{eq}} \subset \chi^{\text{noneq}*} \subset \chi^{\text{noneq}}$ .  
The  $n$ th error of a method  $\bar{X} = \{\bar{X}_n\}_{n \in \mathbb{N}}$  is defined as

$$e_n(\bar{X}) = \|X - \bar{X}_n\|_2 = \left( \mathbb{E} \int_0^T |X(t) - \bar{X}_n(t)|^2 dt \right)^{1/2}. \quad (27)$$

The  $n$ th minimal error, in the respective class of methods under consideration, is defined by

$$e^\diamond(n) = \inf_{\bar{X} \in \chi^\diamond} e_n(\bar{X}), \quad \diamond \in \{\text{eq}, \text{noneq}^*, \text{noneq}\}. \quad (28)$$

We will investigate the exact rate of convergence of the  $n$ th minimal errors (28) together with asymptotic constants. Moreover, we wish to determine (asymptotically) optimal methods  $\bar{X}^\diamond$ ,  $\diamond \in \{\text{eq}, \text{noneq}^*, \text{noneq}\}$ , such that the  $n$ th errors  $e_n(\bar{X}^\diamond)$  tend to zero as fast as  $e^\diamond(n)$  when  $n \rightarrow +\infty$ .

### 3 Asymptotic lower bounds

In this section we investigate asymptotic lower bounds for the problem (1) in the classes of methods  $\chi^\diamond$ ,  $\diamond \in \{\text{eq}, \text{noneq}^*, \text{noneq}\}$ . In the next section we give a construction of approximation methods which are asymptotically optimal. Their definitions will be inspired by the technique used for establishing lower bounds given in this section.

We give the definition of the continuous Milstein approximation and we state its properties that we use in order to establish the lower bounds. Moreover, in next section we use it in order to construct asymptotically optimal methods.

Let  $m \in \mathbb{N}$  and

$$0 = t_0 < t_1 < \dots < t_m = T, \quad (29)$$

be an arbitrary discretization of  $[0, T]$ . We denote by

$$\Delta Z_i = Z(t_{i+1}) - Z(t_i), \quad i = 0, 1, \dots, m-1, \quad (30)$$

for  $Z \in \{N, W\}$ . The continuous Milstein approximation  $\tilde{X}_m^M = \{\tilde{X}_m^M(t)\}_{t \in [0, T]}$  based on (29) is defined as follows. We denote

$$U_i = (t_i, \tilde{X}_m^M(t_i)), \quad (31)$$

and we set

$$\tilde{X}_m^M(0) = x_0, \quad (32)$$

and

$$\begin{aligned} \tilde{X}_m^M(t) &= \tilde{X}_m^M(t_i) + a(U_i) \cdot (t - t_i) + b(U_i) \cdot (W(t) - W(t_i)) + c(U_i) \cdot (N(t) - N(t_i)) \\ &\quad + L_1 b(U_i) \cdot I_{t_i, t}(W, W) + L_{-1} c(U_i) \cdot I_{t_i, t}(N, N) \\ &\quad + L_{-1} b(U_i) \cdot I_{t_i, t}(N, W) + L_1 c(U_i) \cdot I_{t_i, t}(W, N), \end{aligned} \quad (33)$$



for  $t \in [t_i, t_{i+1}]$ ,  $i = 0, 1, \dots, m-1$ , where

$$I_{t_i, t}(Y, Z) = \int_{t_i}^t \int_{t_i}^{s-} dY(u) dZ(s), \quad (34)$$

for  $Y, Z \in \{N, W\}$ . It is well-known that

$$I_{t_i, t}(W, W) = \frac{1}{2} \left( (W(t) - W(t_i))^2 - (t - t_i) \right), \quad (35)$$

$$I_{t_i, t}(N, N) = \frac{1}{2} \left( (N(t) - N(t_i))^2 - (N(t) - N(t_i)) \right), \quad (36)$$

$$I_{t_i, t}(W, N) = \sum_{k=N(t_i)+1}^{N(t)} W(\tau_k) - W(t_i) \cdot (N(t) - N(t_i)), \quad (37)$$

where  $\tau_k$  is the  $k$ th jump time of  $N$ , and

$$I_{t_i, t}(N, W) + I_{t_i, t}(W, N) = (N(t) - N(t_i)) \cdot (W(t) - W(t_i)). \quad (38)$$

Moreover,  $I_{t_i, t}(W, W)$ ,  $I_{t_i, t}(N, N)$ ,  $I_{t_i, t}(N, W)$ ,  $I_{t_i, t}(W, N)$  are independent of  $\mathcal{F}_{t_i}$ , see Fact 6.2 in Appendix.

The main properties of  $\tilde{X}_m^M$  are as follows. For every  $m \in \mathbb{N}$  the process  $\{\tilde{X}_m^M(t)\}_{t \in [0, T]}$  is adapted to  $\{\mathcal{F}_t\}_{t \in [0, T]}$  and has càdlàg paths. The upper bounds on the error of  $\tilde{X}_m^M$  are given in Theorem 6.1. Furthermore, under the commutativity condition (9) the random variables  $\{\tilde{X}_m^M(t_i)\}_{i=0}^m$  are measurable with respect to the sigma field

$$\begin{aligned} \sigma(\mathcal{N}_m(N, W)) &= \sigma(\mathcal{N}_m(N) \oplus \mathcal{N}_m(W)) \\ &:= \sigma(N(t_1), N(t_2), \dots, N(t_m), W(t_1), W(t_2), \dots, W(t_m)). \end{aligned}$$

In particular, this and independence of  $N$  and  $W$  imply that for all  $t \in [t_i, t_{i+1}]$ ,  $i = 0, 1, \dots, m-1$

$$\begin{aligned} \tilde{X}_m^M(t) - \mathbb{E}(\tilde{X}_m^M(t) \mid \mathcal{N}_m(N, W)) &= b(U_i) \cdot (W(t) - \mathbb{E}(W(t) \mid \mathcal{N}_m(W))) \\ &\quad + c(U_i) \cdot (N(t) - \mathbb{E}(N(t) \mid \mathcal{N}_m(N))) \\ &\quad + \tilde{R}_m^M(t), \end{aligned} \quad (39)$$

where

$$\begin{aligned} \tilde{R}_m^M(t) &= L_1 b(U_i) \cdot (I_{t_i, t}(W, W) - \mathbb{E}(I_{t_i, t}(W, W) \mid \mathcal{N}_m(W))) \\ &\quad + L_1 c(U_i) \cdot (I_{t_i, t}(N, W) + I_{t_i, t}(W, N) \\ &\quad \quad - \mathbb{E}(I_{t_i, t}(N, W) + I_{t_i, t}(W, N) \mid \mathcal{N}_m(N, W))) \\ &\quad + L_{-1} c(U_i) \cdot (I_{t_i, t}(N, N) - \mathbb{E}(I_{t_i, t}(N, N) \mid \mathcal{N}_m(N))). \end{aligned} \quad (40)$$

The conditional expectations appearing above can be computed explicitly. Namely, from Lemma 8 in [8] and Lemma 6.2 in Appendix we get by direct calculations

$$\mathbb{E}(W(t) - W(t_i) \mid \mathcal{N}_m(W)) = \Delta W_i \cdot \frac{t - t_i}{t_{i+1} - t_i}, \quad (41)$$

$$\mathbb{E}(N(t) - N(t_i) \mid \mathcal{N}_m(N)) = \Delta N_i \cdot \frac{\Lambda(t, t_i)}{\Lambda(t_{i+1}, t_i)}, \quad (42)$$

$$\mathbb{E}(I_{t_i, t}(W, W) \mid \mathcal{N}_m(W)) = \left( \frac{t - t_i}{t_{i+1} - t_i} \right)^2 \cdot I_{t_i, t_{i+1}}(W, W), \quad (43)$$

$$\mathbb{E}(I_{t_i, t}(N, W) + I_{t_i, t}(W, N) \mid \mathcal{N}_m(N, W)) = \Delta N_i \cdot \frac{\Lambda(t, t_i)}{\Lambda(t_{i+1}, t_i)} \cdot \Delta W_i \cdot \frac{t - t_i}{t_{i+1} - t_i}, \quad (44)$$

$$\mathbb{E}(I_{t_i, t}(N, N) \mid \mathcal{N}_m(N)) = \left( \frac{\Lambda(t, t_i)}{\Lambda(t_{i+1}, t_i)} \right)^2 \cdot I_{t_i, t_{i+1}}(N, N). \quad (45)$$

We stress that for any  $m$  the approximation  $\{\tilde{X}_m^M(t)\}_{t \in [0, T]}$  is not an implementable numerical scheme in our model of computation (even under the commutativity condition (9)), since computation of a trajectory of  $\tilde{X}_m^M$  requires complete knowledge of a corresponding trajectories of  $N$  and  $W$ . However, if the condition (9) holds, by (35), (36) and (38), we can compute values of  $\tilde{X}_m^M$  at the discrete points (29) using only function evaluations of  $W$  and  $N$  at (29).

In order to characterize asymptotic lower bounds we define

$$C^{\text{noneq}} = \frac{1}{\sqrt{6}} \int_0^T \left( \mathbb{E}(\mathcal{Y}(t)) \right)^{1/2} dt, \quad (46)$$

$$C^{\text{eq}} = \sqrt{\frac{T}{6}} \cdot \left( \int_0^T \mathbb{E}(\mathcal{Y}(t)) dt \right)^{1/2}, \quad (47)$$

where the process  $\{\mathcal{Y}(t)\}_{t \in [0, T]}$  is defined in (15). We have that

$$(i) \quad 0 \leq C^{\text{noneq}} \leq C^{\text{eq}},$$

$$(ii) \quad C^{\text{noneq}} = C^{\text{eq}} \text{ iff there exists } \gamma \geq 0 \text{ such that for all } t \in [0, T]$$

$$\mathbb{E}(\mathcal{Y}(t)) = \gamma, \quad (48)$$

$$(iii) \quad C^{\text{eq}} = 0 \text{ iff } C^{\text{noneq}} = 0 \text{ iff } b(t, X(t)) = 0 = c(t, X(t)) \text{ for all } t \in [0, T] \text{ and almost surely.}$$

We have the following result.

**Theorem 3.1** *Let us assume that the mappings  $a$ ,  $b$ ,  $c$  and  $\lambda$  satisfy the assumptions (A)-(E).*

(i) *Let  $\bar{X}$  be an arbitrary method from  $\chi^{\text{noneq}}$ . Then*

$$\liminf_{n \rightarrow +\infty} (\text{cost}_n(\bar{X}))^{1/2} \cdot e_n(\bar{X}) \geq C^{\text{noneq}}. \quad (49)$$

(ii) Let  $\bar{X}$  be an arbitrary method from  $\chi^{\text{noneq}*}$ . If  $b \neq 0$  and  $c \neq 0$  then

$$\liminf_{n \rightarrow +\infty} (\text{cost}_n(\bar{X}))^{1/2} \cdot e_n(\bar{X}) \geq \sqrt{2} \cdot C^{\text{noneq}}. \quad (50)$$

(iii) Let  $\bar{X}$  be an arbitrary method from  $\chi^{\text{eq}}$ . If  $b \neq 0$  and  $c \neq 0$  then

$$\liminf_{n \rightarrow +\infty} (\text{cost}_n(\bar{X}))^{1/2} \cdot e_n(\bar{X}) \geq \sqrt{2} \cdot C^{\text{eq}}, \quad (51)$$

else

$$\liminf_{n \rightarrow +\infty} (\text{cost}_n(\bar{X}))^{1/2} \cdot e_n(\bar{X}) \geq C^{\text{eq}}. \quad (52)$$

**Proof.** We start by showing (49) in the case when  $b \neq 0$  and  $c \neq 0$ . Let  $\bar{X} = \{\bar{X}_n\}_{n \in \mathbb{N}} \in \chi^{\text{noneq}}$  be a method based on an arbitrary sequence of discretizations  $\bar{\Delta}^N = \{\Delta_n^N\}_{n \in \mathbb{N}}$  and  $\bar{\Delta}^W = \{\Delta_n^W\}_{n \in \mathbb{N}}$ , where each  $\Delta_n^N$  and  $\Delta_n^W$  is of the form (19). Every  $\bar{X}_n$  uses information (22) about the processes  $N$  and  $W$ . Take any sequence  $\{m_n\}_{n \in \mathbb{N}}$  of positive integers such that

$$\lim_{n \rightarrow +\infty} \frac{n^{1/2}}{m_n} = \lim_{n \rightarrow +\infty} \frac{m_n}{n} = 0. \quad (53)$$

By  $\hat{\Delta} = \{\hat{\Delta}_n\}_{n \in \mathbb{N}}$  we denote a sequence of discretizations given by  $\{\hat{\Delta}_n\}_{n \in \mathbb{N}} = \{\Delta_n^N \cup \Delta_n^W \cup \Delta_n^{\text{eq}}\}_{n \in \mathbb{N}}$ , where every set  $\Delta_n^{\text{eq}}$  of equidistant points is defined by  $\Delta_n^{\text{eq}} = \{jT/m_n \mid j = 0, 1, \dots, m_n\}$ . Hence, for all  $n \in \mathbb{N}$ ,

$$\hat{\Delta}_n = \{\hat{t}_{0,n}, \hat{t}_{1,n}, \dots, \hat{t}_{k_n,n}\}, \quad (54)$$

with

$$0 = \hat{t}_{0,n} < \hat{t}_{1,n} < \dots < \hat{t}_{k_n,n} = T, \quad \hat{t}_{j,n} \in \Delta_n^N \cup \Delta_n^W \cup \Delta_n^{\text{eq}}, \quad j = 0, 1, \dots, k_n, \quad (55)$$

and

$$n \leq k_n \leq 2n + m_n - 2. \quad (56)$$

Therefore, from (53) and (56) we have that

$$\liminf_{n \rightarrow +\infty} \frac{n}{k_n} \geq \frac{1}{2}, \quad (57)$$

and, since  $\Delta_n^{\text{eq}} \subset \hat{\Delta}_n$  for all  $n \in \mathbb{N}$ ,

$$\max_{0 \leq i \leq k_n-1} (\hat{t}_{i+1,n} - \hat{t}_{i,n}) \leq \frac{T}{m_n}. \quad (58)$$

We denote by  $\hat{\mathcal{N}}(N, W) = \{\hat{\mathcal{N}}_n(N, W)\}_{n \in \mathbb{N}}$ , where each vector  $\hat{\mathcal{N}}_n(N, W)$  consists of the values of  $N$  and  $W$  at  $\hat{\Delta}_n$ , i.e.,

$$\begin{aligned} \hat{\mathcal{N}}_n(N, W) &= \hat{\mathcal{N}}_n(N) \oplus \hat{\mathcal{N}}_n(W) \\ &= [N(\hat{t}_{1,n}), N(\hat{t}_{2,n}), \dots, N(\hat{t}_{k_n,n}), W(\hat{t}_{1,n}), W(\hat{t}_{2,n}), \dots, W(\hat{t}_{k_n,n})]. \end{aligned} \quad (59)$$

Since  $\Delta_n^N \cup \Delta_n^W \subset \hat{\Delta}_n$  for all  $n \in \mathbb{N}$ , we have that

$$\sigma(N(t_{1,n}^N), N(t_{2,n}^N), \dots, N(t_{n,n}^N), W(t_{1,n}^W), W(t_{2,n}^W), \dots, W(t_{n,n}^W)) \subset \sigma(\hat{\mathcal{N}}_n(N, W)). \quad (60)$$

Let us denote by  $\{\tilde{X}_{k_n}^M\}_{n \in \mathbb{N}}$  the sequence of continuous Milstein approximations (32)-(33) based on the sequence of discretizations  $\hat{\Delta}$  and which use the information  $\hat{\mathcal{N}}(N, W)$  about the processes  $N$  and  $W$ . From Theorem 6.1 and (58) we have that

$$\|X - \tilde{X}_{k_n}^M\|_2 \leq C \cdot m_n^{-1}, \quad (61)$$

where the positive constant  $C$  does not depend on  $n$ . Moreover, let

$$\hat{Z}_n(t) = Z(t) - \mathbb{E}(Z(t) \mid \hat{\mathcal{N}}_n(Z)), \quad (62)$$

for  $Z \in \{N, W\}$  and  $t \in [0, T]$ . Note that for any  $t \in [\hat{t}_{i,n}, \hat{t}_{i+1,n}]$  the random variable  $\hat{Z}_n(t)$  is a convex combination of  $Z(t) - Z(\hat{t}_{i,n})$  and  $-(Z(\hat{t}_{i+1,n}) - Z(t))$ . Hence,  $\hat{Z}_n(t)$  is independent of  $\mathcal{F}_{\hat{t}_{i,n}}$  for all  $t \in [\hat{t}_{i,n}, \hat{t}_{i+1,n}]$  and the processes  $\{\hat{N}_n(t)\}_{t \in [0, T]}$ ,  $\{\hat{W}_n(t)\}_{t \in [0, T]}$  are independent. From (58), (60), (61), (39) and Lemma 6.3 we get

$$\begin{aligned} e_n(\bar{X}) &\geq \|\bar{X}_n - \tilde{X}_{k_n}^M\|_2 - \|X - \tilde{X}_{k_n}^M\|_2 \\ &\geq \|\tilde{X}_{k_n}^M - \mathbb{E}(\tilde{X}_{k_n}^M \mid \hat{\mathcal{N}}_n(N, W))\|_2 - C \cdot m_n^{-1} \\ &\geq \left( \sum_{i=0}^{k_n-1} \int_{\hat{t}_{i,n}}^{\hat{t}_{i+1,n}} \mathbb{E}|b(\hat{t}_{i,n}, \tilde{X}_{k_n}^M(\hat{t}_{i,n})) \cdot \hat{W}_n(t) + c(\hat{t}_{i,n}, \tilde{X}_{k_n}^M(\hat{t}_{i,n})) \cdot \hat{N}_n(t)|^2 dt \right)^{1/2} \\ &\quad - \left( \sum_{i=0}^{k_n-1} \int_{\hat{t}_{i,n}}^{\hat{t}_{i+1,n}} \mathbb{E}|\tilde{R}_{k_n}^M(t)|^2 dt \right)^{1/2} - C \cdot m_n^{-1} \\ &\geq \left( \sum_{i=0}^{k_n-1} \int_{\hat{t}_{i,n}}^{\hat{t}_{i+1,n}} \mathbb{E}|b(\hat{t}_{i,n}, \tilde{X}_{k_n}^M(\hat{t}_{i,n}))|^2 \cdot \mathbb{E}|\hat{W}_n(t)|^2 + \mathbb{E}|c(\hat{t}_{i,n}, \tilde{X}_{k_n}^M(\hat{t}_{i,n}))|^2 \cdot \mathbb{E}|\hat{N}_n(t)|^2 dt \right)^{1/2} \\ &\quad - C \cdot m_n^{-1}. \end{aligned} \quad (63)$$

Now, we analyze the asymptotic behavior of the first term in (64). From Lemma 8 in [8] we have that

$$\int_{\hat{t}_{i,n}}^{\hat{t}_{i+1,n}} \mathbb{E}|\hat{W}_n(t)|^2 dt = \frac{1}{6}(\hat{t}_{i+1,n} - \hat{t}_{i,n})^2. \quad (65)$$

For  $i = 0, 1, \dots, k_n - 1$  and  $t \in (\hat{t}_{i,n}, \hat{t}_{i+1,n})$  we define

$$H_{i,n}(t) = \frac{\Lambda(t, \hat{t}_{i,n}) \cdot \Lambda(\hat{t}_{i+1,n}, t)}{(t - \hat{t}_{i,n})(\hat{t}_{i+1,n} - t)}. \quad (66)$$

Of course  $H_{i,n} \in C((\hat{t}_{i,n}, \hat{t}_{i+1,n}))$  and it can be continuously extended to  $[\hat{t}_{i,n}, \hat{t}_{i+1,n}]$ , since  $H(\hat{t}_{i,n}+) = \lambda(\hat{t}_{i,n}) \cdot \Lambda(\hat{t}_{i+1,n}, \hat{t}_{i,n}) / (\hat{t}_{i+1,n} - \hat{t}_{i,n})$  and  $H(\hat{t}_{i+1,n}-) = \lambda(\hat{t}_{i+1,n}) \cdot \Lambda(\hat{t}_{i+1,n}, \hat{t}_{i,n}) / (\hat{t}_{i+1,n} - \hat{t}_{i,n})$  are finite. Therefore, by Lemma 6.2 and from the mean value theorems we get

$$\begin{aligned}
& \int_{\hat{t}_{i,n}}^{\hat{t}_{i+1,n}} \mathbb{E}|\hat{N}_n(t)|^2 dt = \Lambda(\hat{t}_{i+1,n}, \hat{t}_{i,n})^{-1} \cdot \int_{\hat{t}_{i,n}}^{\hat{t}_{i+1,n}} H_{i,n}(t) \cdot (\hat{t}_{i+1,n} - t) \cdot (t - \hat{t}_{i,n}) dt \\
& = \Lambda(\hat{t}_{i+1,n}, \hat{t}_{i,n})^{-1} \cdot H_{i,n}(\hat{d}_{i,n}) \cdot \int_{\hat{t}_{i,n}}^{\hat{t}_{i+1,n}} (\hat{t}_{i+1,n} - t) \cdot (t - \hat{t}_{i,n}) dt \\
& = \frac{1}{6} \frac{\lambda(\hat{\alpha}_{i,n})\lambda(\hat{\beta}_{i,n})}{\lambda(\hat{\gamma}_{i,n})} \cdot (\hat{t}_{i+1,n} - \hat{t}_{i,n})^2 \leq \frac{1}{6} \cdot \|1/\lambda\|_\infty \cdot \|\lambda\|_\infty^2 \cdot (\hat{t}_{i+1,n} - \hat{t}_{i,n})^2, \tag{67}
\end{aligned}$$

for some  $\hat{d}_{i,n}, \hat{\alpha}_{i,n}, \hat{\beta}_{i,n}, \hat{\gamma}_{i,n} \in [\hat{t}_{i,n}, \hat{t}_{i+1,n}]$ ,  $i = 0, 1, \dots, k_n - 1$ . Next, for  $f \in \{b, c\}$  we have from Theorem 6.1 that

$$\begin{aligned}
& \left| \mathbb{E}|f(\hat{t}_{i,n}, \tilde{X}_{k_n}^M(\hat{t}_{i,n}))|^2 - \mathbb{E}|f(\hat{t}_{i,n}, X(\hat{t}_{i,n}))|^2 \right| \leq \\
& \leq C(1 + \sup_{t \in [0, T]} \|\tilde{X}_{k_n}^M(t)\|_{L^2(\Omega)} + \sup_{t \in [0, T]} \|X(t)\|_{L^2(\Omega)}) \cdot \|\tilde{X}_{k_n}^M(\hat{t}_{i,n}) - X(\hat{t}_{i,n})\|_{L^2(\Omega)} \\
& \leq C_1 \max_{0 \leq i \leq k_n - 1} (\hat{t}_{i+1,n} - \hat{t}_{i,n}) \leq \frac{C_1 T}{m_n}. \tag{68}
\end{aligned}$$

Therefore, for  $(f, Z) \in \{(b, W), (c, N)\}$  we have by (65), (67) and (68) that

$$\lim_{n \rightarrow +\infty} n \cdot \sum_{i=0}^{k_n-1} \int_{\hat{t}_{i,n}}^{\hat{t}_{i+1,n}} \left( \mathbb{E}|f(\hat{t}_{i,n}, \tilde{X}_{k_n}^M(\hat{t}_{i,n}))|^2 - \mathbb{E}|f(\hat{t}_{i,n}, X(\hat{t}_{i,n}))|^2 \right) \cdot \mathbb{E}|\hat{Z}_n(t)|^2 dt = 0. \tag{69}$$

This together with the Hölder inequality imply

$$\begin{aligned}
& \liminf_{n \rightarrow +\infty} n \cdot \sum_{i=0}^{k_n-1} \int_{\hat{t}_{i,n}}^{\hat{t}_{i+1,n}} \left( \mathbb{E}|b(\hat{t}_{i,n}, \tilde{X}_{k_n}^M(\hat{t}_{i,n}))|^2 \cdot \mathbb{E}|\hat{W}_n(t)|^2 + \mathbb{E}|c(\hat{t}_{i,n}, \tilde{X}_{k_n}^M(\hat{t}_{i,n}))|^2 \cdot \mathbb{E}|\hat{N}_n(t)|^2 \right) dt \\
& \geq \liminf_{n \rightarrow +\infty} n \cdot \sum_{i=0}^{k_n-1} \int_{\hat{t}_{i,n}}^{\hat{t}_{i+1,n}} \left( \mathbb{E}|b(\hat{t}_{i,n}, X(\hat{t}_{i,n}))|^2 \cdot \mathbb{E}|\hat{W}_n(t)|^2 + \mathbb{E}|c(\hat{t}_{i,n}, X(\hat{t}_{i,n}))|^2 \cdot \mathbb{E}|\hat{N}_n(t)|^2 \right) dt \\
& \geq \liminf_{n \rightarrow +\infty} \frac{n}{6k_n} \cdot \left( \sum_{i=0}^{k_n-1} \left( \mathbb{E}|b(\hat{t}_{i,n}, X(\hat{t}_{i,n}))|^2 + \mathbb{E}|c(\hat{t}_{i,n}, X(\hat{t}_{i,n}))|^2 \cdot \frac{\lambda(\hat{\alpha}_{i,n})\lambda(\hat{\beta}_{i,n})}{\lambda(\hat{\gamma}_{i,n})} \right)^{1/2} \cdot (\hat{t}_{i+1,n} - \hat{t}_{i,n}) \right)^2.
\end{aligned}$$

We have that

$$\begin{aligned}
\hat{S}_n & := \sum_{i=0}^{k_n-1} \left( \mathbb{E}|b(\hat{t}_{i,n}, X(\hat{t}_{i,n}))|^2 + \mathbb{E}|c(\hat{t}_{i,n}, X(\hat{t}_{i,n}))|^2 \cdot \frac{\lambda(\hat{\alpha}_{i,n})\lambda(\hat{\beta}_{i,n})}{\lambda(\hat{\gamma}_{i,n})} \right)^{1/2} \cdot (\hat{t}_{i+1,n} - \hat{t}_{i,n}) \\
& = \hat{S}_n^* + \hat{R}_n^*, \tag{70}
\end{aligned}$$

where

$$\hat{S}_n^* := \sum_{i=0}^{k_n-1} \left( \mathbb{E}(\mathcal{Y}(\hat{t}_{i,n})) \right)^{1/2} \cdot (\hat{t}_{i+1,n} - \hat{t}_{i,n}), \quad (71)$$

and, by (13),

$$\begin{aligned} |\hat{R}_n^*| &= |\hat{S}_n - \hat{S}_n^*| \leq \sum_{i=0}^{k_n-1} \|c(\hat{t}_{i,n}, X(\hat{t}_{i,n}))\|_{L^2(\Omega)} \cdot \left| \frac{\lambda(\hat{\alpha}_{i,n})\lambda(\hat{\beta}_{i,n})}{\lambda(\hat{\gamma}_{i,n})} - \lambda(\hat{t}_{i,n}) \right|^{1/2} \cdot (\hat{t}_{i+1,n} - \hat{t}_{i,n}) \\ &\leq C_1 \cdot \|1/\lambda\|_\infty^{1/2} \cdot \|\lambda\|_\infty^{1/2} \cdot \sum_{i=0}^{k_n-1} \left( |\lambda(\hat{\beta}_{i,n}) - \lambda(\hat{t}_{i,n})|^{1/2} + |\lambda(\hat{\alpha}_{i,n}) - \lambda(\hat{\gamma}_{i,n})|^{1/2} \right) \cdot (\hat{t}_{i+1,n} - \hat{t}_{i,n}) \\ &\leq 2TC_1 \cdot \|1/\lambda\|_\infty^{1/2} \cdot \|\lambda\|_\infty^{1/2} \cdot (\bar{\omega}(\lambda, T/m_n))^{1/2}, \end{aligned} \quad (72)$$

since  $|\sqrt{x} - \sqrt{y}| \leq \sqrt{|x - y|}$  for all  $x, y \geq 0$ . From the uniform continuity of  $\lambda$  we get

$$\lim_{n \rightarrow +\infty} \hat{R}_n^* = 0. \quad (73)$$

Hence, by (70), (71), (73) and Fact 6.1 (ii) we have

$$\lim_{n \rightarrow +\infty} \hat{S}_n = \lim_{n \rightarrow +\infty} \hat{S}_n^* = \int_0^T \left( \mathbb{E}(\mathcal{Y}(t)) \right)^{1/2} dt. \quad (74)$$

Therefore, by (53), (63), (70) and (74) we obtain

$$\liminf_{n \rightarrow +\infty} (cost_n(\bar{X}))^{1/2} \cdot e_n(\bar{X}) \geq \sqrt{2} \cdot \liminf_{n \rightarrow +\infty} \left( \frac{n}{6k_n} \right)^{1/2} \cdot \lim_{n \rightarrow +\infty} \hat{S}_n \geq C^{\text{noneq}}, \quad (75)$$

which ends the proof of (49) in the case when  $b \neq 0$  and  $c \neq 0$ . If  $(b \neq 0 \text{ and } c \equiv 0)$  or  $(b \equiv 0 \text{ and } c \neq 0)$  then  $cost_n(\bar{X}) = n$ ,

$$n \leq k_n \leq n + m_n - 1, \quad (76)$$

and

$$\lim_{n \rightarrow +\infty} \frac{n}{k_n} = 1, \quad (77)$$

which yield

$$\liminf_{n \rightarrow +\infty} (cost_n(\bar{X}))^{1/2} \cdot e_n(\bar{X}) \geq \lim_{n \rightarrow +\infty} \left( \frac{n}{6k_n} \right)^{1/2} \cdot \lim_{n \rightarrow +\infty} \hat{S}_n \geq C^{\text{noneq}}. \quad (78)$$

For  $b \equiv 0$  and  $c \equiv 0$  we obtain trivial lower bound. Finally, if  $\bar{X} \in \chi^{\text{noneq}*}$ ,  $b \neq 0$  and  $c \neq 0$  then  $cost_n(\bar{X}) = 2n$  and by (77) we get

$$\liminf_{n \rightarrow +\infty} (cost_n(\bar{X}))^{1/2} \cdot e_n(\bar{X}) \geq \sqrt{2} \cdot \lim_{n \rightarrow +\infty} \left( \frac{n}{6k_n} \right)^{1/2} \cdot \lim_{n \rightarrow +\infty} \hat{S}_n \geq \sqrt{2} \cdot C^{\text{noneq}}, \quad (79)$$

which completes the proof of (49). The proofs of (52) and (52) are straightforward modifications of the proofs of (49) and (50). Hence, we skip it here. ■

**Remark 3.1** Theorem 3.1 gives nontrivial lower bounds only in the case when

$$\sup_{t \in [0, T]} \mathbb{E}(\mathcal{Y}(t)) > 0. \quad (80)$$

In this case the presented lower bounds still hold even if we allow for methods to use an arbitrary information about  $a$ ,  $b$  and  $c$ , for example, values of partial derivatives or values of arbitrary linear functionals. If  $\mathbb{E}(\mathcal{Y}(t)) = 0$  for all  $t \in [0, T]$  then (1) becomes (almost surely) deterministic ODE. Then different lower bounds hold, see, for example, [13].  $\square$

## 4 Asymptotically optimal methods

We provide definitions of methods that are asymptotically optimal. The construction is inspired by the technique used for establishing the lower bounds in the previous section. We restrict our consideration to approximation methods based on the regular sequences of discretizations generated by a probability density function  $\psi$ , see [28]. For the density  $\psi$  we assume that

(P1)  $\psi \in C([0, T])$  and  $\psi(t) > 0$  for all  $t \in [0, T]$ .

We will use the notation  $\bar{\Delta}_\psi = \{\Delta_{\psi, n}\}_{n \in \mathbb{N}}$  for a sequence of discretizations generated by a density  $\psi$ . The knots

$$\Delta_{\psi, n} = \{t_{0, n}, t_{1, n}, \dots, t_{n, n}\},$$

of the  $n$ th discretization are given by

$$\int_0^{t_{i, n}} \psi(s) ds = \frac{i}{n}, \quad i = 0, 1, \dots, n. \quad (81)$$

Hence, by choosing such a density  $\psi$  one gets a whole sequence of discretizations  $\bar{\Delta}_\psi$ . For instance, the sequence of equidistant discretizations is obtained by taking  $\psi \equiv 1/T$ . Since  $0 < \|\psi\|_\infty^{-1} \leq \|1/\psi\|_\infty < +\infty$ , we have for all  $n \in \mathbb{N}$  and  $i = 0, 1, \dots, n-1$

$$\|\psi\|_\infty^{-1} n^{-1} \leq t_{i+1, n} - t_{i, n} \leq \|1/\psi\|_\infty n^{-1}. \quad (82)$$

We now provide a construction of asymptotically optimal approximation methods. The definition of this method is inspired by the dominating term in the estimation (63). Denote by  $\tilde{X}_\psi^M = \{\tilde{X}_{\psi, n}^M\}_{n \in \mathbb{N}}$  the sequence of continuous Milstein approximations (32)-(33) based on the sequence of discretizations  $\bar{\Delta}_\psi$ . For a given density  $\psi$ , we define the method  $\bar{X}_\psi^{cM} = \{\bar{X}_{\psi, n}^{cM}\}_{n \in \mathbb{N}}$  by

$$\bar{X}_{\psi, n}^{cM}(t) = \mathbb{E}(\tilde{X}_{\psi, n}^M(t) \mid \mathcal{N}_{\psi, n}(N, W)), \quad t \in [0, T], \quad (83)$$

where  $\mathcal{N}_{\psi, n}(N, W)$  consists of values of the processes  $N$  and  $W$  at the points  $\Delta_{\psi, n}$ . (Hence, we formally take  $\bar{\Delta}^W = \bar{\Delta}^N = \bar{\Delta}_\psi$ .) We call (83) *the conditional Milstein method*. We have that  $\bar{X}_{\psi, n}^{cM} \in \chi^{\text{noneq*}}$ . We present an explicit formula for the algorithm (83) in order to show that it

has a form that is allowed in our model of computation. By (9), (33), (83) and (41)-(45) each term  $\bar{X}_{\psi,n}^{cM}$  can be written as

$$\begin{aligned}
\bar{X}_{\psi,n}^{cM}(t) = & \tilde{X}_{\psi,n}^M(t_{i,n}) + a(t_{i,n}, \tilde{X}_{\psi,n}^M(t_{i,n})) \cdot (t - t_{i,n}) \\
& + b(t_{i,n}, \tilde{X}_{\psi,n}^M(t_{i,n})) \cdot \Delta W_{i,n} \cdot \frac{t - t_{i,n}}{t_{i+1,n} - t_{i,n}} \\
& + c(t_{i,n}, \tilde{X}_{\psi,n}^M(t_{i,n})) \cdot \Delta N_{i,n} \cdot \frac{\Lambda(t, t_{i,n})}{\Lambda(t_{i+1,n}, t_{i,n})} \\
& + L_1 b(t_{i,n}, \tilde{X}_{\psi,n}^M(t_{i,n})) \cdot I_{t_{i,n}, t_{i+1,n}}(W, W) \cdot \left( \frac{t - t_{i,n}}{t_{i+1,n} - t_{i,n}} \right)^2 \\
& + L_1 c(t_{i,n}, \tilde{X}_{\psi,n}^M(t_{i,n})) \cdot \Delta N_{i,n} \cdot \Delta W_{i,n} \cdot \frac{\Lambda(t, t_{i,n})}{\Lambda(t_{i+1,n}, t_{i,n})} \cdot \frac{t - t_{i,n}}{t_{i+1,n} - t_{i,n}} \\
& + L_{-1} c(t_{i,n}, \tilde{X}_{\psi,n}^M(t_{i,n})) \cdot I_{t_{i,n}, t_{i+1,n}}(N, N) \cdot \left( \frac{\Lambda(t, t_{i,n})}{\Lambda(t_{i+1,n}, t_{i,n})} \right)^2, \quad (84)
\end{aligned}$$

for  $t \in [t_{i,n}, t_{i+1,n}]$ ,  $i = 0, 1, \dots, n-1$  and  $\bar{X}_{\psi,n}^{cM}(0) = x_0$ . Note that  $\bar{X}_{\psi,n}^{cM}$  has continuous trajectories and coincides with  $\tilde{X}_{\psi,n}^M$  at the discretization points. In general, the method  $\bar{X}_{\psi,n}^{cM}$  is not equal to the piecewise linear interpolation  $\bar{X}_{\psi,n}^{Lin-M}$  of the classical Milstein steps, defined as

$$\bar{X}_{\psi,n}^{Lin-M}(t) = \frac{\tilde{X}_{\psi,n}^M(t_{i,n})(t_{i+1,n} - t) + \tilde{X}_{\psi,n}^M(t_{i+1,n})(t - t_{i,n})}{t_{i+1,n} - t_{i,n}}, \quad (85)$$

for  $t \in [t_{i,n}, t_{i+1,n}]$ ,  $i = 0, 1, \dots, n-1$ , see Remark 4.3. However, we use the method  $\bar{X}_{\psi,n}^{cM}$  in order to investigate the error of  $\bar{X}_{\psi,n}^{Lin-M}$  and we show in the sequel that they behave asymptotically in the same way. Moreover, for a fixed discretization  $\Delta_{\psi,n}$  the method  $\bar{X}_{\psi,n}^{Lin-M}$  does not evaluate  $\Lambda$  and its implementation, at least in the case when  $\psi \equiv 1/T$ , is straightforward.

In the following theorem we give the exact convergence rate of the errors for the methods  $\bar{X}_{\psi}^{cM}$  and  $\bar{X}_{\psi}^{Lin-M}$  in the terms of the following asymptotic constant

$$C_{\psi} = \frac{1}{\sqrt{6}} \left( \int_0^T \frac{\mathbb{E}(\mathcal{Y}(t))}{\psi(t)} dt \right)^{1/2}. \quad (86)$$

The strategy of the proof goes as follows. First, we analyze the error of the conditional Milstein method  $\bar{X}_{\psi}^{cM}$ . Due to its definition given by the conditional expectation (83) this can be done by using some estimates already established in the proof of Theorem 3.1. Then we show that  $\bar{X}_{\psi}^{Lin-M}$  is sufficiently close to  $\bar{X}_{\psi}^{cM}$ . This will give us the asymptotic error for the piecewise linear interpolation method  $\bar{X}_{\psi}^{Lin-M}$ .

**Theorem 4.1** *Let us assume that the mappings  $a, b, c, \lambda$  and  $\psi$  satisfy the assumptions (A)-(E) and (P1), and let  $\bar{X}_{\psi} \in \{\bar{X}_{\psi}^{cM}, \bar{X}_{\psi}^{Lin-M}\}$ . Then if  $b \neq 0$  and  $c \neq 0$*

$$\lim_{n \rightarrow +\infty} (cost_n(\bar{X}_{\psi}))^{1/2} \cdot e_n(\bar{X}_{\psi}) = \sqrt{2} \cdot C_{\psi}, \quad (87)$$



else

$$\lim_{n \rightarrow +\infty} (\text{cost}_n(\bar{X}_\psi))^{1/2} \cdot e_n(\bar{X}_\psi) = C_\psi. \quad (88)$$

**Proof.** From Theorem 6.1 and (82) we get

$$\left| e_n(\bar{X}_\psi^{cM}) - \|\tilde{X}_{\psi,n}^M - \bar{X}_{\psi,n}^{cM}\|_2 \right| \leq e_n(\tilde{X}_\psi^M) \leq C \cdot n^{-1}, \quad (89)$$

where a constant  $C > 0$  does not depend on  $n$ . Moreover, the equality (81) and the integral mean value theorem yield

$$\forall i = 0, 1, \dots, n-1 \exists \xi_{i,n} \in [t_{i,n}, t_{i+1,n}] : n \cdot (t_{i+1,n} - t_{i,n}) = \frac{1}{\psi(\xi_{i,n})}. \quad (90)$$

As in the proof of Theorem 3.1, we use the notation

$$\hat{Z}_{\psi,n}(t) = Z(t) - \mathbb{E}(Z(t) \mid \mathcal{N}_{\psi,n}(Z)), \quad t \in [0, T], \quad (91)$$

for  $Z \in \{N, W\}$ . From (89), (90), Lemma 6.3 and by proceeding analogously as in the proof of Theorem 3.1 we arrive at

$$\begin{aligned} \lim_{n \rightarrow +\infty} n \cdot (e_n(\bar{X}_\psi^{cM}))^2 &= \lim_{n \rightarrow +\infty} n \cdot \|\tilde{X}_{\psi,n}^M - \bar{X}_{\psi,n}^{cM}\|_2^2 \\ &= \lim_{n \rightarrow +\infty} n \cdot \sum_{i=0}^{n-1} \int_{t_{i,n}}^{t_{i+1,n}} \mathbb{E} |b(t_{i,n}, \tilde{X}_{\psi,n}^M(t_{i,n})) \cdot \hat{W}_{\psi,n}(t) + c(t_{i,n}, \tilde{X}_{\psi,n}^M(t_{i,n})) \cdot \hat{N}_{\psi,n}(t)|^2 dt \\ &= \lim_{n \rightarrow +\infty} n \cdot \sum_{i=0}^{n-1} \int_{t_{i,n}}^{t_{i+1,n}} \left( \mathbb{E} |b(t_{i,n}, X(t_{i,n}))|^2 \cdot \mathbb{E} |\hat{W}_{\psi,n}(t)|^2 + \mathbb{E} |c(t_{i,n}, X(t_{i,n}))|^2 \cdot \mathbb{E} |\hat{N}_{\psi,n}(t)|^2 \right) dt \\ &= \frac{1}{6} \lim_{n \rightarrow +\infty} \sum_{i=0}^{n-1} \left( \frac{\mathbb{E} |b(t_{i,n}, X(t_{i,n}))|^2}{\psi(\xi_{i,n})} + \frac{\mathbb{E} |c(t_{i,n}, X(t_{i,n}))|^2}{\psi(\xi_{i,n})} \cdot \frac{\lambda(\alpha_{i,n})\lambda(\beta_{i,n})}{\lambda(\gamma_{i,n})} \right) \cdot (t_{i+1,n} - t_{i,n}) \end{aligned} \quad (92)$$

for some  $d_{i,n}, \alpha_{i,n}, \beta_{i,n}, \gamma_{i,n} \in [t_{i,n}, t_{i+1,n}]$ ,  $i = 0, 1, \dots, n-1$ . Moreover, we have

$$\begin{aligned} &\frac{1}{6} \sum_{i=0}^{n-1} \left( \frac{\mathbb{E} |b(t_{i,n}, X(t_{i,n}))|^2}{\psi(\xi_{i,n})} + \frac{\mathbb{E} |c(t_{i,n}, X(t_{i,n}))|^2}{\psi(\xi_{i,n})} \cdot \frac{\lambda(\alpha_{i,n})\lambda(\beta_{i,n})}{\lambda(\gamma_{i,n})} \right) \cdot (t_{i+1,n} - t_{i,n}) \\ &= \frac{1}{6} \sum_{i=0}^{n-1} \frac{\mathbb{E}(\mathcal{Y}(\xi_{i,n}))}{\psi(\xi_{i,n})} \cdot (t_{i+1,n} - t_{i,n}) + \sum_{i=1}^3 A_{n,i}, \end{aligned} \quad (93)$$

where

$$\begin{aligned} A_{n,1} &= \frac{1}{6} \sum_{i=0}^{n-1} \frac{\mathbb{E} |b(t_{i,n}, X(t_{i,n}))|^2 - \mathbb{E} |b(\xi_{i,n}, X(\xi_{i,n}))|^2}{\psi(\xi_{i,n})} \cdot (t_{i+1,n} - t_{i,n}), \\ A_{n,2} &= \frac{1}{6} \sum_{i=0}^{n-1} \frac{\mathbb{E} |c(t_{i,n}, X(t_{i,n}))|^2 - \mathbb{E} |c(\xi_{i,n}, X(\xi_{i,n}))|^2}{\psi(\xi_{i,n})} \cdot \frac{\lambda(\alpha_{i,n})\lambda(\beta_{i,n})}{\lambda(\gamma_{i,n})} \cdot (t_{i+1,n} - t_{i,n}), \\ A_{n,3} &= \frac{1}{6} \sum_{i=0}^{n-1} \frac{\mathbb{E} |c(\xi_{i,n}, X(\xi_{i,n}))|^2}{\psi(\xi_{i,n})} \cdot \left( \frac{\lambda(\alpha_{i,n})\lambda(\beta_{i,n})}{\lambda(\gamma_{i,n})} - \lambda(\xi_{i,n}) \right) \cdot (t_{i+1,n} - t_{i,n}). \end{aligned}$$

By Fact 6.1 (i), (13) and (82) we get

$$|A_{n,1}| \leq (1/6) \cdot C \cdot \|1/\psi\|_\infty^{3/2} \cdot T \cdot n^{-1/2}, \quad (94)$$

$$|A_{n,2}| \leq (1/6) \cdot C \cdot \|1/\psi\|_\infty^{3/2} \cdot T \cdot \|1/\lambda\|_\infty \cdot \|\lambda\|_\infty^2 \cdot n^{-1/2}, \quad (95)$$

and

$$|A_{n,3}| \leq (T/3) \cdot C \cdot \|1/\psi\|_\infty \cdot \|1/\lambda\|_\infty \cdot \|\lambda\|_\infty \cdot \bar{\omega}(\lambda, \|1/\psi\|_\infty \cdot n^{-1}). \quad (96)$$

This and the uniform continuity of  $\lambda$  imply

$$\lim_{n \rightarrow +\infty} A_{n,j} = 0, \quad j = 1, 2, 3. \quad (97)$$

By (92), (93), (97) and Fact 6.1 (ii) we obtain

$$\lim_{n \rightarrow +\infty} n \cdot (e_n(\bar{X}_\psi^{cM}))^2 = \frac{1}{6} \lim_{n \rightarrow +\infty} \sum_{i=0}^{n-1} \frac{\mathbb{E}(\mathcal{Y}(\xi_{i,n}))}{\psi(\xi_{i,n})} \cdot (t_{i+1,n} - t_{i,n}) = C_\psi^2, \quad (98)$$

which ends the proof of (87) for  $\bar{X}_\psi = \bar{X}_\psi^{cM}$ .

We now analyze the error of  $\bar{X}_{\psi,n}^{Lin-M}$ . Note that

$$\begin{aligned} \bar{R}_{\psi,n}^M(t) &= \bar{X}_{\psi,n}^{cM}(t) - \bar{X}_{\psi,n}^{Lin-M}(t) = c(t_{i,n}, \tilde{X}_{\psi,n}^M(t_{i,n})) \cdot \Delta N_{i,n} \cdot \left( \frac{\Lambda(t, t_{i,n})}{\Lambda(t_{i+1,n}, t_{i,n})} - \frac{t - t_{i,n}}{t_{i+1,n} - t_{i,n}} \right) \\ &+ L_1 b(t_{i,n}, \tilde{X}_{\psi,n}^M(t_{i,n})) \cdot I_{t_{i,n}, t_{i+1,n}}(W, W) \cdot \frac{(t - t_{i,n}) \cdot (t - t_{i+1,n})}{(t_{i+1,n} - t_{i,n})^2} \\ &+ L_1 c(t_{i,n}, \tilde{X}_{\psi,n}^M(t_{i,n})) \cdot \Delta N_{i,n} \cdot \Delta W_{i,n} \cdot \frac{t_{i,n} - t}{t_{i+1,n} - t_{i,n}} \cdot \frac{\Lambda(t_{i+1,n}, t)}{\Lambda(t_{i+1,n}, t_{i,n})} \\ &+ L_{-1} c(t_{i,n}, \tilde{X}_{\psi,n}^M(t_{i,n})) \cdot I_{t_{i,n}, t_{i+1,n}}(N, N) \cdot \left( \left( \frac{\Lambda(t, t_{i,n})}{\Lambda(t_{i+1,n}, t_{i,n})} \right)^2 - \frac{t - t_{i,n}}{t_{i+1,n} - t_{i,n}} \right), \end{aligned}$$

for  $t \in [t_{i,n}, t_{i+1,n}]$ ,  $i = 0, 1, \dots, n-1$ . In addition

$$\begin{aligned} \left| \frac{\Lambda(t, t_{i,n})}{\Lambda(t_{i+1,n}, t_{i,n})} - \frac{t - t_{i,n}}{t_{i+1,n} - t_{i,n}} \right| &\leq C \cdot \left| \lambda(t_{i,n}) \cdot (t_{i+1,n} - t_{i,n}) \cdot \int_{t_{i,n}}^t (\lambda(s) - \lambda(t_{i,n})) ds \right. \\ &\quad \left. + \lambda(t_{i,n}) \cdot (t - t_{i,n}) \cdot \int_{t_{i,n}}^{t_{i+1,n}} (\lambda(t_{i,n}) - \lambda(s)) ds \right| / (t_{i+1,n} - t_{i,n})^2 \\ &\leq C_1 \sup_{t, s \in [t_{i,n}, t_{i+1,n}]} |\lambda(t) - \lambda(s)| \leq C_1 \bar{\omega}(\lambda, \|1/\psi\|_\infty \cdot n^{-1}), \quad (99) \end{aligned}$$

for  $t \in [t_{i,n}, t_{i+1,n}]$ ,  $i = 0, 1, \dots, n-1$ . For  $f \in \{b, c\}$  and  $j \in \{-1, 1\}$  the random variable  $L_j f(t_{i,n}, \tilde{X}_{\psi,n}^M(t_{i,n}))$  is  $\mathcal{F}_{t_{i,n}}$ -measurable and the estimate (180) holds for  $U_i := (t_{i,n}, \tilde{X}_{\psi,n}^M(t_{i,n}))$ .

Hence, it is independent of  $I_{t_{i,n}, t_{i+1,n}}(N, N)$ ,  $I_{t_{i,n}, t_{i+1,n}}(W, W)$  and  $\Delta N_{i,n} \cdot \Delta W_{i,n}$ . Therefore, by (82) we have that

$$\begin{aligned}
\mathbb{E}|\bar{R}_{\psi,n}^M(t)|^2 &\leq 4 \cdot \left( \left| \frac{\Lambda(t, t_{i,n})}{\Lambda(t_{i+1,n}, t_{i,n})} - \frac{t - t_{i,n}}{t_{i+1,n} - t_{i,n}} \right|^2 \cdot \mathbb{E}|c(t_{i,n}, \tilde{X}_{\psi,n}^M(t_{i,n}))|^2 \cdot \mathbb{E}|\Delta N_{i,n}|^2 \right. \\
&\quad + \mathbb{E}|L_1 b(t_{i,n}, \tilde{X}_{\psi,n}^M(t_{i,n}))|^2 \cdot \mathbb{E}|I_{t_{i,n}, t_{i+1,n}}(W, W)|^2 \\
&\quad + \mathbb{E}|L_1 c(t_{i,n}, \tilde{X}_{\psi,n}^M(t_{i,n}))|^2 \cdot \mathbb{E}|\Delta N_{i,n}|^2 \cdot \mathbb{E}|\Delta W_{i,n}|^2 \\
&\quad \left. + \mathbb{E}|L_{-1} c(t_{i,n}, \tilde{X}_{\psi,n}^M(t_{i,n}))|^2 \cdot \mathbb{E}|I_{t_{i,n}, t_{i+1,n}}(N, N)|^2 \right) \\
&\leq C_3 \cdot n^{-1} \cdot (\bar{\omega}(\lambda, \|1/\psi\|_\infty \cdot n^{-1}))^2 + C_4 n^{-2},
\end{aligned} \tag{100}$$

for  $t \in [t_{i,n}, t_{i+1,n}]$ ,  $i = 0, 1, \dots, n-1$ . Since, from (100)

$$\left| e_n(\bar{X}_\psi^{Lin-M}) - e_n(\bar{X}_\psi^{cM}) \right| \leq \|\bar{R}_{\psi,n}^M\|_2 \leq C_4 \cdot n^{-1/2} \cdot \bar{\omega}(\lambda, \|1/\psi\|_\infty \cdot n^{-1}) + C_5 \cdot n^{-1}, \tag{101}$$

we obtain (87) for  $\bar{X}_\psi = \bar{X}_\psi^{Lin-M}$ . This ends the proof.  $\blacksquare$

Let us now assume that the following additional assumption is satisfied:

$$(P2) \quad \inf_{t \in [0, T]} \mathbb{E}(\mathcal{Y}(t)) > 0.$$

The methods  $\bar{X}_\psi^{cM}$  and  $\bar{X}_\psi^{Lin-M}$  obtain the exact rate of convergence  $n^{-1/2}$ , with the asymptotic constant  $C_\psi$  which depends on  $\psi$ . The best density  $\psi_0$ , which is unique and minimizes  $C_\psi$  among

all positive mappings  $\psi \in C([0, T])$  such that  $\int_0^T \psi(t) dt = 1$ , is

$$\psi_0(t) = \frac{\left( \mathbb{E}(\mathcal{Y}(t)) \right)^{1/2}}{\int_0^T \left( \mathbb{E}(\mathcal{Y}(s)) \right)^{1/2} ds}, \quad t \in [0, T]. \tag{102}$$

(The minimization property of  $\psi_0$  follows from the application of the Hölder inequality.) We stress that  $\psi_0$  is strictly positive in  $[0, T]$  under the additional assumption (P2). Furthermore,

$$C_{\psi_0} = C^{\text{noneq}}, \quad \text{and} \quad C_{1/T} = C^{\text{eq}}. \tag{103}$$

The following fact characterizes the case when the equidistant sampling is the optimal one.

**Fact 4.1** *Let us assume that the mappings  $a$ ,  $b$ ,  $c$  and  $\lambda$  satisfy the assumptions (A)-(E) and (P2). Then the following assertions are equivalent.*

(i)  $\psi_0 \equiv 1/T$ .

(ii)  $\mathbb{E}(\mathcal{Y}(t)) = \frac{1}{T} \int_0^T \left( \mathbb{E}(\mathcal{Y}(s)) \right)^{1/2} ds$  for all  $t \in [0, T]$ .

(iii)  $C^{\text{noneq}} = C^{\text{eq}} > 0$ .

**Proof.** The assertion can easily be shown by proving the implications (i)  $\Rightarrow$  (ii)  $\Rightarrow$  (iii)  $\Rightarrow$  (i) and we left it for the reader. ■

From Theorem 4.1 we directly obtain the following result.

**Corollary 4.1** *Let us assume that the mappings  $a$ ,  $b$ ,  $c$  and  $\lambda$  satisfy the assumptions (A)-(E).*

(i) *Let us moreover assume that the assumption (P2) is satisfied. If  $b \not\equiv 0$  and  $c \not\equiv 0$  then for  $\bar{X}_{\psi_0} \in \{\bar{X}_{\psi_0}^{cM}, \bar{X}_{\psi_0}^{Lin-M}\}$  it holds*

$$\lim_{n \rightarrow +\infty} (\text{cost}_n(\bar{X}_{\psi_0}))^{1/2} \cdot e_n(\bar{X}_{\psi_0}) = \sqrt{2} \cdot C^{\text{noneq}}, \quad (104)$$

*else*

$$\lim_{n \rightarrow +\infty} (\text{cost}_n(\bar{X}_{\psi_0}))^{1/2} \cdot e_n(\bar{X}_{\psi_0}) = C^{\text{noneq}}. \quad (105)$$

(ii) *Let  $\bar{X}_{1/T} \in \{\bar{X}_{1/T}^{cM}, \bar{X}_{1/T}^{Lin-M}\}$ . If  $b \not\equiv 0$  and  $c \not\equiv 0$  then it holds*

$$\lim_{n \rightarrow +\infty} (\text{cost}_n(\bar{X}_{1/T}))^{1/2} \cdot e_n(\bar{X}_{1/T}) = \sqrt{2} \cdot C^{\text{eq}}, \quad (106)$$

*else*

$$\lim_{n \rightarrow +\infty} (\text{cost}_n(\bar{X}_{1/T}))^{1/2} \cdot e_n(\bar{X}_{1/T}) = C^{\text{eq}}. \quad (107)$$

Theorem 3.1 and Corollary 4.1 imply the main result of the paper.

**Theorem 4.2** *Let us assume that the mappings  $a$ ,  $b$ ,  $c$  and  $\lambda$  satisfy the assumptions (A)-(E).*

(i) *Let us additionally assume that the assumption (P2) is satisfied. If ( $b \not\equiv 0$  and  $c \equiv 0$ ) or ( $b \equiv 0$  and  $c \not\equiv 0$ ) then*

$$\lim_{n \rightarrow +\infty} n^{1/2} \cdot e^{\text{noneq}}(n) = C^{\text{noneq}}, \quad (108)$$

*and the methods  $\bar{X}_{\psi_0}^{cM}, \bar{X}_{\psi_0}^{Lin-M}$ , where  $\psi_0$  is defined in (102), are asymptotically optimal in the class  $\chi^{\text{noneq}}$ . If  $b \not\equiv 0$  and  $c \not\equiv 0$  then*

$$C^{\text{noneq}}/\sqrt{2} \leq \liminf_{n \rightarrow +\infty} n^{1/2} \cdot e^{\text{noneq}}(n) \leq \limsup_{n \rightarrow +\infty} n^{1/2} \cdot e^{\text{noneq}}(n) \leq C^{\text{noneq}}. \quad (109)$$

(ii) *If the assumption (P2) is satisfied,  $b \not\equiv 0$  and  $c \not\equiv 0$  then*

$$\lim_{n \rightarrow +\infty} n^{1/2} \cdot e^{\text{noneq}*}(n) = C^{\text{noneq}},$$

*and the methods  $\bar{X}_{\psi_0}^{cM}, \bar{X}_{\psi_0}^{Lin-M}$  are asymptotically optimal in the class  $\chi^{\text{noneq}*}$ .*

(iii) *We have that*

$$\lim_{n \rightarrow +\infty} n^{1/2} \cdot e^{\text{eq}}(n) = C^{\text{eq}},$$

*and the methods  $\bar{X}_{1/T}^{cM}, \bar{X}_{1/T}^{Lin-M}$  are asymptotically optimal in the class  $\chi^{\text{eq}}$ .*

As we can see the optimal rate of convergence of the minimal errors in the classes  $\chi^{\text{eq}}$  and  $\chi^{\text{noneq}}$  is proportional to  $n^{-1/2}$ , where  $n$  is a total number of evaluations of  $N$  and  $W$ . In the class  $\chi^{\text{noneq}}$  we have a gap between upper and lower asymptotic constants. We conjecture that (108) holds also if  $b \neq 0$  and  $c \neq 0$ .

We end this section with the following remarks.

**Remark 4.1** Theorem 4.2 implies that the error can be reduced asymptotically by the factor

$$C^{\text{eq}}/C^{\text{noneq}},$$

if we use the optimal discretization instead of the equidistant one. However, the optimal density  $\psi_0$  and the optimal sampling  $\{t_{i,n}\}_{i=0}^n$ , defined by

$$\int_0^{t_{i,n}} \left( \mathbb{E}(\mathcal{Y}(t)) \right)^{1/2} dt = \frac{i}{n} \int_0^T \left( \mathbb{E}(\mathcal{Y}(t)) \right)^{1/2} dt, \quad i = 0, 1, \dots, n, \quad (110)$$

can be computed explicitly only in particular cases see, for example, Section 4.1. Moreover, the additional assumption (P2) is required. We plan to overwhelm these difficulties in the future work.  $\square$

**Remark 4.2** If  $c \equiv 0$ ,  $b \neq 0$  and  $T = 1$  then, for the classes  $\chi^{\text{eq}}$  and  $\chi^{\text{noneq}}$ , Theorem 4.2 restores the results of Theorem 2 (iii) and Proposition 2 from [12] in the Gaussian case, while if  $c = c(t)$ ,  $b \equiv 0$  and  $\lambda = \text{const}$  then we get Theorem 4.2 from [26] for the pure jump case with an additive Poisson noise. In addition to this paper, in [26] the author established a method based on an adaptive stepsize control that does not depend on the knowledge of  $\lambda$ . The problem of defining such methods for SDEs of the general type (1) will be the topic of our future work.  $\square$

**Remark 4.3** We have  $\bar{X}_\psi^{cM} \equiv \bar{X}_\psi^{Lin-M}$ , if  $\lambda = \text{const}$ ,  $b = b(t)$  and  $c = c(t)$ .  $\square$

#### 4.1 Linear case - Merton's jump diffusion model

Let us consider the following SDE

$$\begin{cases} dX(t) = rX(t)dt + \sigma X(t)dW(t) + X(t-)dN(t), & t \in [0, T], \\ X(0) = x_0 > 0, \end{cases} \quad (111)$$

that models the stock price in the Merton's model, see [21]. We assume  $\lambda$  to be a constant function and  $r \in \mathbb{R}$ ,  $\sigma > 0$ . The solution of (111) is

$$X(t) = x_0 \exp\left((r - \frac{1}{2}\sigma^2)t + \sigma W(t)\right) \cdot 2^{N(t)}. \quad (112)$$

We denote  $\gamma = r + \sigma^2/2 + 3\lambda/2$  and we have

$$\mathbb{E}(\mathcal{Y}(t)) = (\sigma^2 + \lambda) \cdot x_0^2 \cdot e^{2\gamma t} > 0, \quad \text{for all } t \in [0, T]. \quad (113)$$

If  $\gamma = 0$  then the optimal sampling is the equidistant one and  $C^{\text{noneq}} = C^{\text{eq}} = Tx_0\sqrt{\frac{\sigma^2+\lambda}{6}}$ . If  $\gamma \neq 0$  then we obtain the following optimal sampling for (111)

$$t_{i,n} = \frac{1}{\gamma} \ln \left( \frac{i}{n} (e^{\gamma T} - 1) + 1 \right), \quad t = 0, 1, \dots, n. \quad (114)$$

We have that  $t_{i,n} \rightarrow iT/n$  for  $\gamma \rightarrow 0$ . Since  $C^{\text{noneq}}/C^{\text{eq}}$  behaves as  $\sqrt{2/\gamma T}$  when  $\gamma T \rightarrow +\infty$ , we can gain by using the nonequidistant mesh.

## 5 Conclusions

We investigated the minimal asymptotic errors for strong global approximation of SDEs driven by the Poisson and Wiener processes. We considered the cases of equidistant and nonequidistant sampling of  $N$  and  $W$ . In both cases, we showed that the minimal error tends to zero like  $Cn^{-1/2}$ , where  $C$  is an average in time of a local Hölder constant of  $X$  and  $n$  is the number of evaluations of  $N$  and  $W$ . However, the asymptotic constant  $C$  in the case of equidistant sampling can be considerably larger than the asymptotic constant when nonuniform mesh is used. We provided a construction of methods that asymptotically achieve the established minimal errors.

In this paper, we addressed the case when sampling points for the processes  $N$  and  $W$  are chosen only in the nonadaptive way with respect to  $N$  and  $W$ . Moreover, we assume that the diffusion and jump coefficients satisfied the jump commutativity condition. For the adaptive sampling and non-commutative case preliminary considerations indicate that the direct application of methods developed in this paper is not possible. Further extension of the presented analysis is needed in that case and we postpone this problem to our future work.

## Acknowledgments

Part of this work was done at Banff International Research Station for Mathematical Innovation and Discovery (BIRS), Alberta, Canada, where the author participated at the workshop "Approximation of High-Dimensional Numerical Problems - Algorithms, Analysis and Applications", Fall 2015. I would like to thank the Staff of the BIRS for great hospitality.

## 6 Appendix

We use the following version of the Itô formula for semimartingales with jumps, see, for example, [29] or [22].

**Lemma 6.1** *Let us assume that the mappings  $a$ ,  $b$ ,  $c$  and  $\lambda$  satisfy the assumptions (B1), (B2)*

and (E). Let a function  $U : \mathbb{R} \rightarrow \mathbb{R}$  belongs to  $C^2(\mathbb{R})$ . Then for the solution  $X$  of (1) it holds

$$\begin{aligned} U(X(t)) = U(X(0)) &+ \int_0^t \left( U'(X(s)) \cdot a(s, X(s)) + \frac{1}{2} \cdot U''(X(s)) \cdot b^2(s, X(s)) \right) ds \\ &+ \int_0^t U'(X(s)) \cdot b(s, X(s)) dW(s) \\ &+ \int_0^t \left( U(X(s-)) + c(s, X(s-)) - U(X(s-)) \right) dN(s). \end{aligned}$$

The proof of the following fact is straightforward.

**Fact 6.1** *Let the mappings  $a, b, c$  and  $\lambda$  satisfy the assumptions (B1), (B2) and (E).*

(i) *There exists a constant  $C_1 > 0$  such that for all  $f \in \{b, c\}$  and  $t, s \in [0, T]$  we have*

$$\left| \mathbb{E}|f(t, X(t))|^2 - \mathbb{E}|f(s, X(s))|^2 \right| \leq C_1 |t - s|^{1/2}. \quad (115)$$

(ii) *The mapping*

$$[0, T] \ni t \rightarrow \mathbb{E}(\mathcal{Y}(t)) \in \mathbb{R}_+ \cup \{0\}, \quad (116)$$

*is continuous.*

(iii) *There exists a constant  $C_2 > 0$  such that*

$$\mathbb{E} \left( \sup_{t \in [0, T]} \mathcal{Y}(t) \right) \leq C_2. \quad (117)$$

**Fact 6.2** (i) *There exists  $C > 0$  such that for all  $0 \leq s \leq t \leq T$  and  $Y, Z \in \{N, W\}$  we have*

$$\mathbb{E}|I_{s,t}(Y, Z)|^2 \leq C(t - s)^2. \quad (118)$$

(ii) *For all  $0 \leq s \leq t \leq T$  and  $Y, Z \in \{N, W\}$  the stochastic integral  $I_{s,t}(Y, Z)$  is independent of  $\mathcal{F}_s$ .*

**Proof.** The proof of (i) can be straightforwardly delivered from (6), (38), the isometry for stochastic integrals driven by martingales and by the independence of  $W$  and  $N$ . Hence, we skip it.

For the proof of (ii) note that directly from (35) and (36) we get that  $I_{s,t}(Y, Y)$ ,  $Y \in \{N, W\}$ , is independent of  $\mathcal{F}_s$ . So the only case of interest is when  $(Y, Z) \in \{(N, W), (W, N)\}$ .

Fix  $s, t \in [0, T]$ ,  $s \leq t$ , and let  $\Delta_m = \{\alpha_{0,m}, \alpha_{1,m}, \dots, \alpha_{m,m}\}$ ,  $m \in \mathbb{N}$ , be a sequence of discretizations of  $[s, t]$  such that  $s = \alpha_{0,m} < \alpha_{1,m} < \dots < \alpha_{m,m} = t$  and  $\lim_{m \rightarrow +\infty} \|\Delta_m\| = 0$ , where  $\|\Delta_m\| = \max_{0 \leq i \leq m-1} (\alpha_{i+1,m} - \alpha_{i,m})$ . Moreover, let

$$I_{s,t}^m(N, W) = \sum_{j=0}^{m-1} (N(\alpha_{j,m}) - N(s)) \cdot (W(\alpha_{j+1,m}) - W(\alpha_{j,m})). \quad (119)$$

We have that

$$I_{s,t}(N, W) = \lim_{m \rightarrow +\infty} I_{s,t}^m(N, W) \quad \text{in } L^2(\Omega). \quad (120)$$

Therefore, the sequence  $\{I_{s,t}^m(N, W)\}_{m \in \mathbb{N}}$  converges also in probability and, by the independence of the increments of  $N$  and  $W$ , every random variable  $I_{s,t}^m(N, W)$  is independent of  $\mathcal{F}_s$ . Hence, the limit  $I_{s,t}(N, W)$  is also independent of  $\mathcal{F}_s$ . By (38) we get that also  $I_{s,t}(W, N)$  is independent of  $\mathcal{F}_s$ . ■

**The proof of Proposition 2.1.** By the Markov property of the solution  $X$  we have that  $\|X(t+h) - X(t) \mid X(t)\|_{L^2(\Omega)} = \|X(t+h) - X(t) \mid \mathcal{F}_t\|_{L^2(\Omega)}$ . For all  $t \in [0, T)$  and  $h > 0$  such that  $0 \leq t < t+h \leq T$  we have

$$\begin{aligned} X(t+h) - X(t) &= \int_t^{t+h} \left( a(s, X(s)) + \lambda(s)c(s, X(s)) \right) ds \\ &\quad + \int_t^{t+h} b(s, X(s)) dW(s) + \int_t^{t+h} c(s, X(s-)) d\tilde{N}(s). \end{aligned} \quad (121)$$

From (10) and (E) we obtain that

$$\mathbb{E} \left( \left| \int_t^{t+h} \left( a(s, X(s)) + \lambda(s)c(s, X(s)) \right) ds \right|^2 \mid \mathcal{F}_t \right) \leq C_1 \cdot h^2 + C_1 \cdot h \cdot \mathbb{E} \left( \int_t^{t+h} |X(s)|^2 ds \mid \mathcal{F}_t \right), \quad (122)$$

almost surely. By Theorem 88 in [29] we obtain for all  $t \in [0, T)$  and almost surely

$$I_1(W) := \mathbb{E} \left( \left| \int_t^{t+h} b(s, X(s)) dW(s) \right|^2 \mid \mathcal{F}_t \right) = \mathbb{E} \left( \int_t^{t+h} |b(s, X(s))|^2 ds \mid \mathcal{F}_t \right), \quad (123)$$

$$I_2(N) := \mathbb{E} \left( \left| \int_t^{t+h} c(s, X(s-)) d\tilde{N}(s) \right|^2 \mid \mathcal{F}_t \right) = \mathbb{E} \left( \int_t^{t+h} \lambda(s) \cdot |c(s, X(s))|^2 ds \mid \mathcal{F}_t \right), \quad (124)$$

and

$$\begin{aligned} &\mathbb{E} \left( \int_t^{t+h} b(s, X(s)) dW(s) \cdot \int_t^{t+h} c(s, X(s-)) d\tilde{N}(s) \mid \mathcal{F}_t \right) \\ &= \mathbb{E} \left( \int_t^{t+h} b(s, X(s)) \cdot c(s, X(s-)) d\langle W, \tilde{N} \rangle(s) \mid \mathcal{F}_t \right) = 0, \end{aligned} \quad (125)$$



since  $(W(t) \cdot \tilde{N}(t), \mathcal{F}_t)_{t \in [0, T]}$  is a martingale. Therefore, by Minkowski's inequality for conditional expectations (see [5]), we have that

$$\begin{aligned} & \frac{1}{h^{1/2}} \left( I_1(W) + I_2(N) \right)^{1/2} - \left( C_1 \cdot h + C_1 \cdot \mathbb{E} \left( \int_t^{t+h} |X(s)|^2 ds \mid \mathcal{F}_t \right) \right)^{1/2} \\ & \leq \frac{\|X(t+h) - X(t) \mid \mathcal{F}_t\|_{L^2(\Omega)}}{h^{1/2}} \\ & \leq \frac{1}{h^{1/2}} \left( I_1(W) + I_2(N) \right)^{1/2} + \left( C_1 \cdot h + C_1 \cdot \mathbb{E} \left( \int_t^{t+h} |X(s)|^2 ds \mid \mathcal{F}_t \right) \right)^{1/2}, \end{aligned} \quad (126)$$

almost surely. From (13), Fact 6.1 (iii) and the Lebesgue's dominated convergence theorem for conditional expectations (see [5]) we have for all  $t \in [0, T)$  and almost surely that

$$\lim_{h \rightarrow 0+} \mathbb{E} \left( \int_t^{t+h} |X(s)|^2 ds \mid \mathcal{F}_t \right) = 0, \quad (127)$$

and

$$\lim_{h \rightarrow 0+} \frac{I_1(W) + I_2(N)}{h} = \lim_{h \rightarrow 0+} \mathbb{E} \left( \frac{1}{h} \int_t^{t+h} \mathcal{Y}(s) ds \mid \mathcal{F}_t \right) = \mathcal{Y}(t), \quad (128)$$

since  $X$  and  $\mathcal{Y}$  have càdlàg paths and  $\mathcal{Y}(t)$  is  $\mathcal{F}_t$ -measurable. This together with (126) yield (16). Now, (17) follows from (16) and Lebesgue's dominated convergence theorem. ■

**Lemma 6.2** *Let  $m \in \mathbb{N}$  and let*

$$0 = t_0 < t_1 < \dots < t_m = T, \quad (129)$$

*be an arbitrary discretization of the interval  $[0, T]$  and*

$$\mathcal{N}_m(N) = [N(t_1), N(t_2), \dots, N(t_m)]. \quad (130)$$

*Then for all  $i = 0, 1, \dots, m-1$  and  $t \in [t_i, t_{i+1}]$*

*(i)*

$$\mathbb{E}(N(t) \mid \mathcal{N}_m(N)) = \frac{N(t_{i+1}) \cdot \Lambda(t, t_i) + N(t_i) \cdot \Lambda(t_{i+1}, t)}{\Lambda(t_{i+1}, t_i)}, \quad (131)$$

*almost surely,*

*(ii)*

$$\mathbb{E} \left( |N(t) - \mathbb{E}(N(t) \mid \mathcal{N}_m(N))|^2 \mid \mathcal{N}_m(N) \right) = (N(t_{i+1}) - N(t_i)) \cdot \frac{\Lambda(t_{i+1}, t) \cdot \Lambda(t, t_i)}{(\Lambda(t_{i+1}, t_i))^2}, \quad (132)$$

*almost surely and, in particular,*

$$\mathbb{E} |N(t) - \mathbb{E}(N(t) \mid \mathcal{N}_m(N))|^2 = \frac{\Lambda(t_{i+1}, t) \cdot \Lambda(t, t_i)}{\Lambda(t_{i+1}, t_i)}. \quad (133)$$

**Proof.** For  $t = t_i$ ,  $i = 0, 1, \dots, m$ , we directly get (131), (132) and (133). By the results of [2], from the fact that the process  $N$  has independent increments and by direct calculations we obtain that conditioned on  $\mathcal{N}_m(N)$  and for  $t \in (t_i, t_{i+1})$ ,  $i = 0, 1, \dots, m-1$  the increment  $N(t) - N(t_i)$  is a binomial random variable with the number of trials  $N(t_{i+1}) - N(t_i)$  and with the probability of success in each trial equal to  $\frac{\Lambda(t, t_i)}{\Lambda(t_{i+1}, t_i)}$ . Now, the rest of proof goes analogously as the proof of Lemma 3.1 in [26]. ■

We provide a result concerning an upper bound on the error for the continuous Milstein approximation  $\tilde{X}_m^M$ . A similar result has been shown in Theorem 6.4.1 in [21], however, under slightly stronger assumptions. In particular, we do not assume the existence of continuous partial derivative  $\partial f / \partial t$  for  $f \in \{a, b, c\}$  and we do not impose here any Lipschitz conditions on the second partial derivative of  $f = f(t, y)$ ,  $f \in \{a, b, c\}$ , with respect to  $y$ . Moreover, we consider here nonstationary Poisson process, while in [21] Theorem 6.4.1 has been proven for stationary point processes.

**Theorem 6.1** *Let us assume that the mappings  $a, b, c$  and  $\lambda$  satisfy the assumptions (A), (B), (C) and (E). Let  $m \in \mathbb{N}$  and let (29) be an arbitrary discretization of the interval  $[0, T]$ . Then for the continuous Milstein approximation  $\tilde{X}_m^M$ , based on the mesh (29), we have that*

$$\sup_{t \in [0, T]} \|\tilde{X}_m^M(t)\|_{L^2(\Omega)} \leq C_1, \quad (134)$$

and

$$\sup_{t \in [0, T]} \|X(t) - \tilde{X}_m^M(t)\|_{L^2(\Omega)} \leq C_2 \cdot \max_{0 \leq i \leq m-1} (t_{i+1} - t_i), \quad (135)$$

where  $C_1, C_2 > 0$  do not depend on  $m$ .

**Proof.** Recall that  $U_i = (t_i, \tilde{X}_m^M(t_i))$  and  $L_j f(U_i)$  is  $\mathcal{F}_{t_i}$ -measurable for  $f \in \{b, c\}$ ,  $j \in \{-1, 1\}$ . First, we show that

$$\sup_{t \in [0, T]} \|\tilde{X}_m^M(t)\|_{L^2(\Omega)} < +\infty. \quad (136)$$

We proceed by induction. Let us assume that  $\|\tilde{X}_m^M(t_k)\|_{L^2(\Omega)} < +\infty$  for  $k = 0, 1, \dots, i$  and some  $i$ . (The assumption is fulfilled for  $i = 0$ .) By (10), (12), (33) and Fact 6.2 we have for all  $k = 0, 1, \dots, i$  and  $t \in [t_k, t_{k+1}]$  that

$$\|\tilde{X}_m^M(t) - \tilde{X}_m^M(t_k)\|_{L^2(\Omega)} \leq C(1 + \|\tilde{X}_m^M(t_k)\|_{L^2(\Omega)})(t - t_k)^{1/2}. \quad (137)$$

Hence,  $\sup_{t \in [t_i, t_{i+1}]} \|\tilde{X}_m^M(t)\|_{L^2(\Omega)} < +\infty$  and, in particular,  $\|\tilde{X}_m^M(t_{i+1})\|_{L^2(\Omega)} < +\infty$ . Therefore, we

get  $\max_{i=0,1,\dots,n} \|\tilde{X}_m^M(t_i)\|_{L^2(\Omega)} < +\infty$  and (136).

We now justify (135). The solution  $X = X(t)$  of (1) and the continuous Milstein approximation  $\tilde{X}_m^M = \tilde{X}_m^M(t)$  can be written as

$$X(t) = x_0 + A(t) + B(t) + C(t), \quad (138)$$

$$\tilde{X}_m^M(t) = x_0 + \tilde{A}_m^M(t) + \tilde{B}_m^M(t) + \tilde{C}_m^M(t), \quad (139)$$

where

$$A(t) = \int_0^t \sum_{i=0}^{m-1} a(s, X(s)) \cdot \mathbf{1}_{(t_i, t_{i+1}]}(s) ds, \quad (140)$$

$$B(t) = \int_0^t \sum_{i=0}^{m-1} b(s, X(s)) \cdot \mathbf{1}_{(t_i, t_{i+1}]}(s) dW(s), \quad (141)$$

$$C(t) = \int_0^t \sum_{i=0}^{m-1} c(s, X(s-)) \cdot \mathbf{1}_{(t_i, t_{i+1}]}(s) dN(s), \quad (142)$$

and

$$\begin{aligned} \tilde{A}_m^M(t) &= \int_0^t \sum_{i=0}^{m-1} a(U_i) \cdot \mathbf{1}_{(t_i, t_{i+1}]}(s) ds, \\ \tilde{B}_m^M(t) &= \int_0^t \sum_{i=0}^{m-1} \left( b(U_i) + \int_{t_i}^s L_1 b(U_i) dW(u) + \int_{t_i}^s L_{-1} b(U_i) dN(u) \right) \cdot \mathbf{1}_{(t_i, t_{i+1}]}(s) dW(s), \\ \tilde{C}_m^M(t) &= \int_0^t \sum_{i=0}^{m-1} \left( c(U_i) + \int_{t_i}^s L_1 c(U_i) dW(u) + \int_{t_i}^{s-} L_{-1} c(U_i) dN(u) \right) \cdot \mathbf{1}_{(t_i, t_{i+1}]}(s) dN(s). \end{aligned}$$

We have for all  $t \in [0, T]$  that

$$\mathbb{E}|A(t) - \tilde{A}_m^M(t)|^2 \leq 3 \left( \mathbb{E}|\tilde{A}_{m,1}^M(t)|^2 + \mathbb{E}|\tilde{A}_{m,2}^M(t)|^2 + \mathbb{E}|\tilde{A}_{m,3}^M(t)|^2 \right), \quad (143)$$

where

$$\mathbb{E}|\tilde{A}_{m,1}^M(t)|^2 = \mathbb{E} \left| \int_0^t \sum_{i=0}^{m-1} \left( a(s, X(s)) - a(t_i, X(s)) \right) \cdot \mathbf{1}_{(t_i, t_{i+1}]}(s) ds \right|^2, \quad (144)$$

$$\mathbb{E}|\tilde{A}_{m,2}^M(t)|^2 = \mathbb{E} \left| \int_0^t \sum_{i=0}^{m-1} \left( a(t_i, X(s)) - a(t_i, X(t_i)) \right) \cdot \mathbf{1}_{(t_i, t_{i+1}]}(s) ds \right|^2, \quad (145)$$

$$\mathbb{E}|\tilde{A}_{m,3}^M(t)|^2 = \mathbb{E} \left| \int_0^t \sum_{i=0}^{m-1} \left( a(t_i, X(t_i)) - a(t_i, \tilde{X}_m^M(t_i)) \right) \cdot \mathbf{1}_{(t_i, t_{i+1}]}(s) ds \right|^2. \quad (146)$$

We get from the Hölder inequality and Lemma 2.1 for all  $t \in [0, T]$  that

$$\mathbb{E}|\tilde{A}_{m,1}^M(t)|^2 \leq TK^2 \sum_{i=0}^{m-1} \int_{t_i}^{t_{i+1}} (s - t_i)^2 \cdot \mathbb{E}(1 + |X(s)|)^2 ds \leq C \max_{0 \leq i \leq m-1} (t_{i+1} - t_i)^2. \quad (147)$$

From Lemma 6.1 applied to  $U(x) = a(t_i, x)$  and (6) we have that for  $s \in [t_i, t_{i+1}]$

$$\begin{aligned}
a(t_i, X(s)) - a(t_i, X(t_i)) &= \int_{t_i}^s \left( \frac{\partial a}{\partial y}(t_i, X(u)) \cdot a(u, X(u)) + \frac{1}{2} \frac{\partial^2 a}{\partial y^2}(t_i, X(u)) \cdot b^2(u, X(u)) \right) du \\
&\quad + \int_{t_i}^s \frac{\partial a}{\partial y}(t_i, X(u)) \cdot b(u, X(u)) dW(u) \\
&\quad + \int_{t_i}^s \left( a(t_i, X(u-) + c(u, X(u-))) - a(t_i, X(u-)) \right) d\tilde{N}(u) \\
&\quad + \int_{t_i}^s \left( a(t_i, X(u-) + c(u, X(u-))) - a(t_i, X(u-)) \right) \lambda(u) du.
\end{aligned}$$

We denote for  $f \in \{a, b, c\}$  and  $u \in (t_i, t_{i+1}]$

$$\begin{aligned}
\alpha_i(f, u) &= \frac{\partial f}{\partial y}(t_i, X(u)) \cdot a(u, X(u)) + \frac{1}{2} \frac{\partial^2 f}{\partial y^2}(t_i, X(u)) \cdot b^2(u, X(u)), \\
\beta_i(f, u) &= \frac{\partial f}{\partial y}(t_i, X(u)) \cdot b(u, X(u)), \\
\gamma_i(f, u) &= f(t_i, X(u-) + c(u, X(u-))) - f(t_i, X(u-)).
\end{aligned}$$

We have

$$\mathbb{E}|\tilde{A}_{m,2}^M(t)|^2 \leq 4 \left( \mathbb{E}|\tilde{M}_{m,1}^M(t)|^2 + \mathbb{E}|\tilde{M}_{m,2}^M(t)|^2 + \mathbb{E}|\tilde{M}_{m,3}^M(t)|^2 + \mathbb{E}|\tilde{M}_{m,4}^M(t)|^2 \right), \quad (148)$$

where

$$\mathbb{E}|\tilde{M}_{m,1}^M(t)|^2 = \mathbb{E} \left| \int_0^t \sum_{i=0}^{m-1} \left( \int_{t_i}^s \alpha_i(a, u) du \right) \cdot \mathbf{1}_{(t_i, t_{i+1}]}(s) ds \right|^2, \quad (149)$$

$$\mathbb{E}|\tilde{M}_{m,2}^M(t)|^2 = \mathbb{E} \left| \int_0^t \sum_{i=0}^{m-1} \left( \int_{t_i}^s \beta_i(a, u) dW(u) \right) \cdot \mathbf{1}_{(t_i, t_{i+1}]}(s) ds \right|^2, \quad (150)$$

$$\mathbb{E}|\tilde{M}_{m,3}^M(t)|^2 = \mathbb{E} \left| \int_0^t \sum_{i=0}^{m-1} \left( \int_{t_i}^s \gamma_i(a, u) d\tilde{N}(u) \right) \cdot \mathbf{1}_{(t_i, t_{i+1}]}(s) ds \right|^2, \quad (151)$$

$$\mathbb{E}|\tilde{M}_{m,4}^M(t)|^2 = \mathbb{E} \left| \int_0^t \sum_{i=0}^{m-1} \left( \int_{t_i}^s \gamma_i(a, u) \cdot \lambda(u) du \right) \cdot \mathbf{1}_{(t_i, t_{i+1}]}(s) ds \right|^2, \quad (152)$$

for all  $t \in [0, T]$ . By the Hölder inequality, (10), (11) and Lemma 2.1 we have

$$\mathbb{E}|\tilde{M}_{m,1}^M(t)|^2 \leq C_1 \sum_{i=0}^{m-1} \int_{t_i}^{t_{i+1}} \left( (s - t_i) \int_{t_i}^s \mathbb{E}(1 + |X(u)|)^4 du \right) ds \leq C_3 \max_{0 \leq i \leq m-1} (t_{i+1} - t_i)^2. \quad (153)$$

By Theorem 6.5.8 in [17] and Theorem 88 (iii) in [29] we obtain for all  $t \in [0, T]$

$$\begin{aligned}\mathbb{E}|\tilde{M}_{m,2}^M(t)|^2 &= \int_0^t \int_0^t \sum_{i=0}^{m-1} \mathbb{E} \left( \int_{t_i}^{\min\{s_1, s_2\}} |\beta_i(a, u)|^2 du \right) \cdot \mathbf{1}_{(t_i, t_{i+1}]}(s_1) \cdot \mathbf{1}_{(t_i, t_{i+1}]}(s_2) ds_1 ds_2 \\ &\leq C \max_{0 \leq i \leq m-1} (t_{i+1} - t_i)^2,\end{aligned}\tag{154}$$

$$\begin{aligned}\mathbb{E}|\tilde{M}_{m,3}^M(t)|^2 &= \int_0^t \int_0^t \sum_{i=0}^{m-1} \mathbb{E} \left( \int_{t_i}^{\min\{s_1, s_2\}} |\gamma_i(a, u)|^2 \cdot \lambda(u) du \right) \cdot \mathbf{1}_{(t_i, t_{i+1}]}(s_1) \cdot \mathbf{1}_{(t_i, t_{i+1}]}(s_2) ds_1 ds_2 \\ &\leq C \max_{0 \leq i \leq m-1} (t_{i+1} - t_i)^2.\end{aligned}\tag{155}$$

We estimate (151) analogously as (149) and we get for all  $t \in [0, T]$  that

$$\mathbb{E}|\tilde{M}_{m,4}^M(t)|^2 \leq C \max_{0 \leq i \leq m-1} (t_{i+1} - t_i)^2.\tag{156}$$

Hence, by (148), (153), (154) and (156) we arrive at

$$\mathbb{E}|\tilde{A}_{m,2}^M(t)|^2 \leq C \max_{0 \leq i \leq m-1} (t_{i+1} - t_i)^2,\tag{157}$$

for all  $t \in [0, T]$ . For (146) we have by the Hölder inequality and (A2) that

$$\mathbb{E}|\tilde{A}_{m,3}^M(t)|^2 \leq \lambda^2 K T \int_0^t \sum_{i=0}^{m-1} \mathbb{E}|X(t_i) - \tilde{X}_m^M(t_i)|^2 \cdot \mathbf{1}_{(t_i, t_{i+1}]}(s) ds,\tag{158}$$

for all  $t \in [0, T]$ . Hence, (143), (147), (157) and (158) yield for all  $t \in [0, T]$  that

$$\mathbb{E}|A(t) - \tilde{A}_m^M(t)|^2 \leq C_1 \int_0^t \sum_{i=0}^{m-1} \mathbb{E}|X(t_i) - \tilde{X}_m^M(t_i)|^2 \cdot \mathbf{1}_{(t_i, t_{i+1}]}(s) ds + C_2 \max_{0 \leq i \leq m-1} (t_{i+1} - t_i)^2.\tag{159}$$

We have for all  $t \in [0, T]$  that

$$\mathbb{E}|B(t) - \tilde{B}_m^M(t)|^2 \leq 3 \left( \mathbb{E}|\tilde{B}_{m,1}^M(t)|^2 + \mathbb{E}|\tilde{B}_{m,2}^M(t)|^2 + \mathbb{E}|\tilde{B}_{m,3}^M(t)|^2 \right),\tag{160}$$

where

$$\begin{aligned}\mathbb{E}|\tilde{B}_{m,1}^M(t)|^2 &= \mathbb{E} \left| \int_0^t \sum_{i=0}^{m-1} \left( b(s, X(s)) - b(t_i, X(s)) \right) \cdot \mathbf{1}_{(t_i, t_{i+1}]}(s) dW(s) \right|^2, \\ \mathbb{E}|\tilde{B}_{m,2}^M(t)|^2 &= \mathbb{E} \left| \int_0^t \sum_{i=0}^{m-1} \left( b(t_i, X(s)) - b(t_i, X(t_i)) \right. \right. \\ &\quad \left. \left. - \int_{t_i}^s L_1 b(U_i) dW(u) - \int_{t_i}^s L_{-1} b(U_i) dN(u) \right) \cdot \mathbf{1}_{(t_i, t_{i+1}]}(s) dW(s) \right|^2, \\ \mathbb{E}|\tilde{B}_{m,3}^M(t)|^2 &= \mathbb{E} \left| \int_0^t \sum_{i=0}^{m-1} \left( b(t_i, X(t_i)) - b(U_i) \right) \cdot \mathbf{1}_{(t_i, t_{i+1}]}(s) dW(s) \right|^2.\end{aligned}$$

From the Itô isometry and the Hölder inequality we obtain for  $t \in [0, T]$

$$\mathbb{E}|\tilde{B}_{m,1}^M(t)|^2 \leq \sum_{i=0}^{m-1} \int_{t_i}^{t_{i+1}} \mathbb{E}|b(s, X(s)) - b(t_i, X(s))|^2 ds \leq C \max_{0 \leq i \leq m-1} (t_{i+1} - t_i)^2. \quad (161)$$

By the Itô isometry together with the Itô formula we get

$$\mathbb{E}|\tilde{B}_{m,2}^M(t)|^2 \leq 4 \left( \mathbb{E}|\bar{M}_{m,1}^M(t)|^2 + \mathbb{E}|\bar{M}_{m,2}^M(t)|^2 + \mathbb{E}|\bar{M}_{m,3}^M(t)|^2 + \mathbb{E}|\bar{M}_{m,4}^M(t)|^2 \right), \quad (162)$$

where

$$\mathbb{E}|\bar{M}_{m,1}^M(t)|^2 = \mathbb{E} \int_0^t \sum_{i=0}^{m-1} \left( \int_{t_i}^s \alpha_i(b, u) du \right)^2 \cdot \mathbf{1}_{(t_i, t_{i+1}]}(s) ds, \quad (163)$$

$$\mathbb{E}|\bar{M}_{m,2}^M(t)|^2 = \mathbb{E} \int_0^t \sum_{i=0}^{m-1} \left( \int_{t_i}^s (\beta_i(b, u) - L_1 b(U_i)) dW(u) \right)^2 \cdot \mathbf{1}_{(t_i, t_{i+1}]}(s) ds, \quad (164)$$

$$\mathbb{E}|\bar{M}_{m,3}^M(t)|^2 = \mathbb{E} \int_0^t \sum_{i=0}^{m-1} \left( \int_{t_i}^s (\gamma_i(b, u) - L_{-1} b(U_i)) d\tilde{N}(u) \right)^2 \cdot \mathbf{1}_{(t_i, t_{i+1}]}(s) ds, \quad (165)$$

$$\mathbb{E}|\bar{M}_{m,4}^M(t)|^2 = \mathbb{E} \int_0^t \sum_{i=0}^{m-1} \left( \int_{t_i}^s (\gamma_i(b, u) - L_{-1} b(U_i)) \cdot \lambda(u) du \right)^2 \cdot \mathbf{1}_{(t_i, t_{i+1}]}(s) ds, \quad (166)$$

for all  $t \in [0, T]$ . From the Hölder inequality we get

$$\mathbb{E}|\bar{M}_{m,1}^M(t)|^2 \leq C(1 + \sup_{t \in [0, T]} \mathbb{E}|X(t)|^4) \max_{0 \leq i \leq m-1} (t_{i+1} - t_i)^2, \quad (167)$$

for all  $t \in [0, T]$ . Since we have for  $u \in [t_i, t_{i+1}]$  that

$$\mathbb{E}|\beta_i(b, u) - L_1 b(U_i)|^2 \leq C \left( (1 + \sup_{t \in [0, T]} \mathbb{E}|X(t)|^2) \cdot (u - t_i)^2 + (u - t_i) + \mathbb{E}|X(t_i) - \tilde{X}_m^M(t_i)|^2 \right), \quad (168)$$

we obtain

$$\mathbb{E}|\bar{M}_{m,2}^M(t)|^2 \leq C_1 \int_0^t \sum_{i=0}^{m-1} \mathbb{E}|X(t_i) - \tilde{X}_m^M(t_i)|^2 \cdot \mathbf{1}_{(t_i, t_{i+1}]}(s) ds + C_2 \max_{0 \leq i \leq m-1} (t_{i+1} - t_i)^2. \quad (169)$$

Moreover, for  $s \in [t_i, t_{i+1}]$  we have that

$$\mathbb{E} \int_{t_i}^s |\gamma_i(b, u) - L_{-1} b(U_i)|^2 du \leq K_1 \cdot \mathbb{E}|X(t_i) - \tilde{X}_m^M(t_i)|^2 + K_2 \cdot (t_{i+1} - t_i)^2, \quad (170)$$

which implies

$$\begin{aligned}\mathbb{E}|\bar{M}_{m,3}^M(t)|^2 &= \int_0^t \sum_{i=0}^{m-1} \left( \int_{t_i}^s \mathbb{E}|\gamma_i(b, u) - L_{-1}b(U_i)|^2 \cdot \lambda(u) du \right) \cdot \mathbf{1}_{(t_i, t_{i+1}]}(s) ds \\ &\leq C_1 \int_0^t \sum_{i=0}^{m-1} \mathbb{E}|X(t_i) - \tilde{X}_m^M(t_i)|^2 \cdot \mathbf{1}_{(t_i, t_{i+1}]}(s) ds + C_2 \max_{0 \leq i \leq m-1} (t_{i+1} - t_i)^2,\end{aligned}\quad (171)$$

$$\begin{aligned}\mathbb{E}|\bar{M}_{m,4}^M(t)|^2 &= \int_0^t \sum_{i=0}^{m-1} \left( (s - t_i) \cdot \int_{t_i}^s \mathbb{E}|\gamma_i(b, u) - L_{-1}b(U_i)|^2 \cdot (\lambda(u))^2 du \right) \cdot \mathbf{1}_{(t_i, t_{i+1}]}(s) ds \\ &\leq C_1 \int_0^t \sum_{i=0}^{m-1} \mathbb{E}|X(t_i) - \tilde{X}_m^M(t_i)|^2 \cdot \mathbf{1}_{(t_i, t_{i+1}]}(s) ds + C_2 \max_{0 \leq i \leq m-1} (t_{i+1} - t_i)^3.\end{aligned}\quad (172)$$

Hence, for all  $t \in [0, T]$  we get

$$\mathbb{E}|\tilde{B}_{m,2}^M(t)|^2 \leq K^2 \int_0^t \sum_{i=0}^{m-1} \mathbb{E}|X(t_i) - \tilde{X}_m^M(t_i)|^2 \cdot \mathbf{1}_{(t_i, t_{i+1}]}(s) ds + C_2 \max_{0 \leq i \leq m-1} (t_{i+1} - t_i)^2, \quad (173)$$

$$\mathbb{E}|\tilde{B}_{m,3}^M(t)|^2 \leq K^2 \int_0^t \sum_{i=0}^{m-1} \mathbb{E}|X(t_i) - \tilde{X}_m^M(t_i)|^2 \cdot \mathbf{1}_{(t_i, t_{i+1}]}(s) ds. \quad (174)$$

Therefore, by (160), (161), (173) and (174) we obtain for all  $t \in [0, T]$

$$\mathbb{E}|B(t) - \tilde{B}_m^M(t)|^2 \leq C_1 \int_0^t \sum_{i=0}^{m-1} \mathbb{E}|X(t_i) - \tilde{X}_m^M(t_i)|^2 \cdot \mathbf{1}_{(t_i, t_{i+1}]}(s) ds + C_2 \max_{0 \leq i \leq m-1} (t_{i+1} - t_i)^2. \quad (175)$$

Now

$$\mathbb{E}|C(t) - \tilde{C}_m^M(t)|^2 \leq 3 \left( \mathbb{E}|\tilde{C}_{m,1}^M(t)|^2 + \mathbb{E}|\tilde{C}_{m,2}^M(t)|^2 + \mathbb{E}|\tilde{C}_{m,3}^M(t)|^2 \right), \quad (176)$$

where

$$\begin{aligned}\mathbb{E}|\tilde{C}_{m,1}^M(t)|^2 &= \mathbb{E} \left| \int_0^t \sum_{i=0}^{m-1} \left( c(s, X(s-)) - c(t_i, X(s-)) \right) \cdot \mathbf{1}_{(t_i, t_{i+1}]}(s) dN(s) \right|^2, \\ \mathbb{E}|\tilde{C}_{m,2}^M(t)|^2 &= \mathbb{E} \left| \int_0^t \sum_{i=0}^{m-1} \left( c(t_i, X(s-)) - c(t_i, X(t_i)) \right. \right. \\ &\quad \left. \left. - \int_{t_i}^s L_1 c(U_i) dW(u) - \int_{t_i}^{s-} L_{-1} c(U_i) dN(u) \right) \cdot \mathbf{1}_{(t_i, t_{i+1}]}(s) dN(s) \right|^2, \\ \mathbb{E}|\tilde{C}_{m,3}^M(t)|^2 &= \mathbb{E} \left| \int_0^t \sum_{i=0}^{m-1} \left( c(t_i, X(t_i)) - c(U_i) \right) \cdot \mathbf{1}_{(t_i, t_{i+1}]}(s) dN(s) \right|^2,\end{aligned}$$

for all  $t \in [0, T]$ . Next, we use the decomposition  $dN(t) = d\tilde{N}(t) + dm(t) = d\tilde{N}(t) + \lambda(t)dt$  and the martingale isometry. Then the estimation of the above terms goes in analogous way as for  $\mathbb{E}|B(t) - \tilde{B}_m^M(t)|^2$ , hence, we skip it. We get for all  $t \in [0, T]$  that

$$\mathbb{E}|C(t) - \tilde{C}_m^M(t)|^2 \leq C_1 \int_0^t \sum_{i=0}^{m-1} \mathbb{E}|X(t_i) - \tilde{X}_m^M(t_i)|^2 \cdot \mathbf{1}_{(t_i, t_{i+1}]}(s) ds + C_2 \max_{0 \leq i \leq m-1} (t_{i+1} - t_i)^2. \quad (177)$$

Combining (159), (175) and (177) we get for all  $t \in [0, T]$

$$\sup_{0 \leq s \leq t} \mathbb{E}|X(s) - \tilde{X}_m^M(s)|^2 \leq C_1 \int_0^t \sup_{0 \leq u \leq s} \mathbb{E}|X(u) - \tilde{X}_m^M(u)|^2 ds + C_2 \max_{0 \leq i \leq m-1} (t_{i+1} - t_i)^2. \quad (178)$$

By (13) and (136) the mapping  $[0, T] \ni t \rightarrow \sup_{0 \leq s \leq t} \mathbb{E}|X(s) - \tilde{X}_m^M(s)|^2 \in \mathbb{R}_+ \cup \{0\}$  is bounded and Borel measurable. Hence, by Gronwall's lemma we get (135). The estimate (134) is a consequence of (13) and (135). This ends the proof. ■

**Lemma 6.3** *Let us assume that the mappings  $a, b, c$  and  $\lambda$  satisfy the assumptions (A)-(E). For all  $t \in [t_i, t_{i+1}]$ ,  $i = 0, 1, \dots, m-1$*

$$\mathbb{E}|\tilde{R}_m^M(t)|^2 \leq C(t_{i+1} - t_i)^2, \quad (179)$$

where  $C > 0$  does not depend on  $m$  nor  $i$ .

**Proof.** From (12) and Theorem 6.1 we have for  $f \in \{b, c\}$  and  $j \in \{-1, 1\}$  that

$$\mathbb{E}|L_j f(U_i)|^2 \leq C, \quad (180)$$

where  $C > 0$  does not depend on  $m$  nor  $i$ . Moreover, for  $f \in \{b, c\}$  and  $j \in \{-1, 1\}$  the random variable  $L_j f(U_i)$  is  $\mathcal{F}_{t_i}$ -measurable. From Fact 6.2 (ii) and by (41)-(45) we have that  $I_{t_i, t}(N, N) - \mathbb{E}(I_{t_i, t}(N, N) \mid \mathcal{N}_m(N))$ ,  $I_{t_i, t}(W, W) - \mathbb{E}(I_{t_i, t}(W, W) \mid \mathcal{N}_m(W))$  and  $I_{t_i, t}(N, W) + I_{t_i, t}(W, N) - \mathbb{E}(I_{t_i, t}(N, W) + I_{t_i, t}(W, N) \mid \mathcal{N}_m(N, W))$  are independent of  $\mathcal{F}_{t_i}$ . Hence, by (40), Fact 6.2 (i) and (180) we get

$$\begin{aligned} \|\tilde{R}_m^M(t)\|_{L^2(\Omega)} &\leq 2 \cdot \|L_1 b(U_i)\|_{L^2(\Omega)} \cdot \|I_{t_i, t}(W, W)\|_{L^2(\Omega)} + 2 \cdot \|L_{-1} c(U_i)\|_{L^2(\Omega)} \cdot \|I_{t_i, t}(N, N)\|_{L^2(\Omega)} \\ &\quad + 2 \cdot \|L_1 c(U_i)\|_{L^2(\Omega)} \cdot \|I_{t_i, t}(N, W) + I_{t_i, t}(W, N)\|_{L^2(\Omega)} \\ &\leq C(t - t_i), \end{aligned} \quad (181)$$

for  $t \in [t_i, t_{i+1}]$ , which ends the proof of (179). ■

## 7 References

- [1] APPLEBAUM, D., *Lévy Processes and Stochastic Calculus*. 2nd ed., Cambridge University Press, 2011.



- [2] BONET, E., NUALART, D., Interpolation and forecasting in Poisson's processes. *Stochastica* **2** (1977), 1–5.
- [3] BRUTI-LIBERATI, N., PLATEN, E., Strong approximations of stochastic differential equations with jumps. *J. Comput. Appl. Math.* **205** (2007), 982–1001.
- [4] DĘBOWSKI, J., PRZYBYŁOWICZ, P., Optimal approximation of stochastic integrals with respect to a homogeneous Poisson process, submitted.
- [5] DOOB, J. L., *Measure Theory*. Springer-Verlag New York, 1994.
- [6] GARDÓN, A., The order of approximations for solutions of Itô-type stochastic differential equations with jumps. *Stoch. Anal. Appl.* **22** (2004), 679–699.
- [7] GRAHAM, C., TALAY, D., *Stochastic Simulation and Monte Carlo Methods*. Springer-Verlag, Berlin, Heidelberg, 2013.
- [8] HERTLING, P., Nonlinear Lebesgue and Itô integration problems of high complexity. *J. Complexity* **17** (2001), 366–387.
- [9] HIGHAM, D.J., KLOEDEN, P.E., Numerical methods for nonlinear stochastic differential equations with jumps. *Numer. Math.* **101** (2005), 101–119.
- [10] HIGHAM, D.J., KLOEDEN, P.E., Convergence and stability of implicit methods for jump-diffusion systems. *Int. J. Numer. Anal. Model.* **3** (2006), 125–140.
- [11] HIGHAM, D.J., KLOEDEN, P.E., Strong convergence rates for backward Euler on a class of nonlinear jump-diffusion problems. *J. Comput. Appl. Math.* **205** (2007), 949–956.
- [12] HOFMANN, N., MÜLLER-GRONBACH, T., RITTER, K., The optimal discretization of stochastic differential equations. *J. Complexity* **17** (2001), 117–153.
- [13] KACEWICZ, B., Minimal asymptotic error of algorithms for solving ODE. *J. Complexity* **4** (1988), 373–389.
- [14] KLOEDEN, P.E., PLATEN, E., *Numerical Solution of Stochastic Differential Equations*. Springer Verlag, Berlin, Heidelberg, 1992.
- [15] KOPP, P.E., *Martingales and Stochastic Integrals*. Cambridge University Press, 1984.
- [16] KUMAR, C., SABANIS, S., On tamed Milstein schemes of SDEs driven by Lévy noise, to appear in *Discr. and Cont. Dyn. Sys. - Series B*, see also <http://arxiv.org/abs/1407.5347>
- [17] KUO, H-H., *Introduction to Stochastic Integration*. Springer Science, 2006.
- [18] MÜLLER-GRONBACH, T., Strong approximation of systems of stochastic differential equations. *Habilitationsschrift*, TU Darmstadt (2002).

- [19] NEUENKIRCH, A., Optimal approximation of SDE's with additive fractional noise. *J. Complexity* **22** (2006), 459–474.
- [20] NEUENKIRCH, A., *Optimal approximation of stochastic differential equations with additive fractional noise*. Ph.D. Thesis, TU Darmstadt. Shaker Verlag, Aachen, 2006.
- [21] PLATEN, E., BRUTI-LIBERATI, N., *Numerical Solution of Stochastic Differential Equations with Jumps in Finance*. Springer Verlag, Berlin, Heidelberg, 2010.
- [22] PROTTER, P. E., *Stochastic Integration and Differential Equations*. 2nd ed., Springer-Verlag Berlin Heidelberg, 2004.
- [23] PRZYBYŁOWICZ, P., MORKISZ, P., Strong approximation of solutions of stochastic differential equations with time-irregular coefficients via randomized Euler algorithm. *Appl. Numer. Math.* **78** (2014), 80–94.
- [24] PRZYBYŁOWICZ, P., Minimal asymptotic error for one-point approximation of SDEs with time-irregular coefficients. *J. Comp. Appl. Math.* **282** (2015), 98–110.
- [25] PRZYBYŁOWICZ, P., Optimal global approximation of SDEs with time-irregular coefficients in asymptotic setting. *Appl. Math. Comp.* **270** (2015), 441–457.
- [26] PRZYBYŁOWICZ, P., Optimal global approximation of stochastic differential equations with additive Poisson noise. *Numer. Algor.* **73** (2016), 323–348.
- [27] RITTER, K., *Average Case Analysis of Numerical Problems*. Lecture Notes in Math. **1733**, Springer Verlag, Berlin, 2000.
- [28] SACKS, J., YLVISAKER, D., Design for regression problems with correlated errors III, *Ann. Math. Stat.* **41** (1970), 2057–2074.
- [29] SITU, R., *Theory of Stochastic Differential Equations with Jumps and Applications*. Springer Science, 2005.
- [30] SOBCZYK, K., *Stochastic Differential Equations with Applications to Physics and Engineering*. Kluwer Academic Publishers B.V., 1991.
- [31] TRAUB, J.F., WASILKOWSKI G.W., WOŹNIAKOWSKI, H., *Information-Based Complexity*. Academic Press, New York, 1988.